ABSTRACT

Title of dissertation: EVALUATING *IN VITRO* AND *IN VIVO* LOOPING EFFICIENCIES OF ARTIFICIAL DNA-BINDING PROTEINS

Sarah Mae C. Sucayan
Doctor of Philosophy, 2014

Dissertation directed by: Associate Professor Jason D. Kahn
Department of Chemistry and Biochemistry

DNA looping plays an important role in gene regulation by increasing the local concentration of a transcriptional activator or repressor at its primary binding site. Several *in vitro* and *in vivo* studies on DNA looping showed that the stability of protein-mediated DNA loops depends on the flexibility of both the looping protein and the DNA that contains the binding sites. We designed two types of short and rigid DNA looping proteins, based on a coiled-coil motif, in order to probe DNA flexibility on the thermodynamics of protein-mediated DNA loop formation. *In vitro* characterization of the putatively tetrameric DNA binding protein lzee by electrophoretic mobility shift assays (EMSA) did not show evidence of a sandwich complex, which is a necessary precursor to DNA looping. A quantitative *in vivo* looping assay, adapted from the reporter gene assay developed by Becker, Kahn, and Maher (2005), showed relatively weak enhancement of repression on GCN4 operators spaced >300 bp apart by lzee and the dimeric looping proteins LZD73, LZD80 and LZD87. However, lzee and LZD87 expression triggers cell toxicity or

highly decreased reporter protein expression on reporter strains containing GCN4 sites <240 bp apart. We proposed recombination events to explain the unexpected behavior in this distance range. Results from *in vitro* Plasmid Conformation Capture (PCC) revealed a very weak increase in crosslinking efficiency on 450- and 900-bp DNA loops. The apparent failure in capturing DNA loops by *in vivo* and *in vitro* PCC was attributed to the LZD proteins not being able to crosslink to DNA. Lastly, we introduced two kinds of modifications to our DNA binding proteins. The first modification sought to improve the linker sequence in lzee in order to select for better tetrameric looping proteins. The other modification introduced lysine residues at the DNA binding domains in the dimeric GCN4 looping protein LZD87 to enhance their ability to crosslink DNA. The *in vivo* repression assay failed to select for lzee mutants that are better repressors than lzee, while the crosslinking assays on the LZD single mutants did not show clear evidence that the new proteins can crosslink DNA. Taking all of these results together, we have concluded that the inability of the LZD proteins to stabilize DNA loops *in vivo* support the model that DNA plays a more passive role in the thermodynamics of DNA looping. However, the unique ability of lzee to trigger recombination in our repression assays can be utilized to design assays that detect recombination as a consequence of DNA looping, although further studies are required to understand this behavior. The molecular tools presented in this work would serve not only in providing a deeper understanding the thermodynamics of DNA looping, but could also be used as a starting point to develop better systems for modulating gene expression.

EVALUATING *IN VITRO* AND *IN VIVO* LOOPING EFFICIENCIES OF ARTIFICIAL DNA-BINDING PROTEINS

by

Sarah Mae Cruzada Sucayan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Associate Professor Jason D. Kahn, Chair/Advisor
Professor David Fushman
Associate Professor Douglas A. Julin
Associate Professor Stephen M. Mount
Professor William E. Bentley, Dean's Representative

Dedication

Para ini sa akon ginikanan, mga kauturan, kag sa akon mga kahinablusan. Labi sa tanan, para man ini sa akon hinigugma, sa iya nga walay kahimoan nga pagsakdag sang akon mga handum.

Acknowledgments

My time in graduate school has been long and arduous, but it has been a life-changing experience for me. For that, I am forever grateful to the following people..

To Dr. Jason D. Kahn, my thesis advisor and mentor.. for welcoming me into his lab, for letting me take the reins to my project, and for providing me with the guidance when I needed it. Thank you for giving me the opportunity to grow professionally, and for reminding me to always "scan the entire forest and not get hung up on the individual leaves."

The members of my Ph. D. Committee.. for their invaluable criticisms and advice as they guide me through the process of pursuing a doctorate degree.

All the members of the Kahn Lab.. for you helpful comments and suggestions during lab presentations. I am especially thankful for Aaron, Kathy, and Dan – lab members who started graduate school with me.. for being more than just colleagues, you have been instrumental for me to finally adapt to the rigors and challenges of graduate school.

My past and present housemates and roommates, who have hailed from all walks of life.. for teaching me to keep an open mind on everything. For most of you, you have been more than just people who I live with. I am very thankful for your friendship.

Friends from back home and all over the world who I have kept in touch, and friends who have introduced me to activities and experiences that I have never

thought possible.. for making my life more meaningful. It's been quite an adventure so far. Thank you for all of the wonderful memories.

My parents and my siblings.. for their unwavering love and support in all of my endeavours, even if means we have to be physically apart for a very long time. They have always been my #1 cheering squad.

Finally, to Marti, my beloved.. for being the most stable and calming presence that I can always rely on. I thank God to have met you at just the right time, and I am very lucky to have you in my life. Обичам те от тук до небето и още три пъти по толкова!

Contents

List of Tables

List of Figures

List of Abbreviations

3C	chromosome conformation capture
4HB	four helix bundle (tetramerization domain)
6x His	6x Histidine (affinity tag)
aa	amino acid
amp	ampicillin
AP-1	activator protein 1
ara	arabinose
Bluo-gal	5-bromo-3-indolyl-β-D-galactopyranoside
cam	chloramphenicol
cAMP	cyclic adenosine monophosphate
CAP	catabolite activator protein
ChIP	chromatin immunoprecipitation
CREB	cAMP response element-binding protein
CV	column volume
DBD	DNA binding domain
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
FPLC	fast protein liquid chromatography
gen	gentamicin
GST	glutathione-S-transferase (affinity tag)
Gua-HCl	guanidine hydrochloride
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	isopropyl-β-D-thiogalactopyranoside
kan	kanamycin
lacA	galactoside-O-acetyltransferase (protein)
lacI	lactose repressor (protein)
lacY	lactose permease (protein)
lacZ	β-galactosidase (protein)
LB	Luria-Bertani
LZD	leucine zipper dual-binding (protein)
LZIP	leucine zipper (dimerization domain)
LZT	leucine zipper tetrameric (protein)
MCS	multiple cloning site
MES	2-(N-morpholino)ethanesulfonic acid
ONPG	o-nitrophenyl-β-galactosidase
PAA	proline-alanine-alanine (GCN4 protein dimer)

PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCC	plasmid conformation capture
PCR	polymerase chain reaction
PNK	polynucleotide kinase
RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
ssDNA	single-stranded DNA
strep	streptomycin
tet	tetracycline
TBE	tris, borate, and EDTA (buffer)
TBS	tris-buffered saline
TBST	tris-buffered saline with Tween 20
Tris	tris(hydroxymethyl)aminomethane
TTC	2,3,5-triphenyl-2H-tetrazolium chloride
TZ-glu	tetrazolium-glucose
TZ-lac	tetrazolium-lactose
WLC	Worm-Like Chain
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
IDT	Integrated DNA Technologies
NEB	New England Biolabs

Chapter 1: Introduction

1.1 On DNA looping and gene regulation

Gene regulation is an essential process for the growth, development, and evolution of many organisms. A variety of genes would be expressed at different times, places, and in different combinations during the lifetime of an organism. In *Escherichia coli*, not all genes in its chromosome are expressed at the same time. Rather, signals from the cell's environment dictate which of these genes get transcribed. For more complex organisms, the intricate regulation of gene expression in different cells would result to different products of these genes, i.e. enzymes and structural proteins, that determine the cell architecture and behavior in the different parts of the organism.

There are two general steps in gene expression that are common in all organisms. First, the gene is transcribed into messenger RNA (mRNA), and second, the mRNA gets translated into protein. More complex organisms have developed additional steps in controlling gene expression. However, the most common way to regulate gene expression, which is the least taxing to the cell, is to control the transcription initiation (Matthews, 1992). Given the limited amounts of transcriptional machinery in a cell at any given time, organisms have developed a wide array of tran-

scription factors. These are DNA binding proteins whose sole purpose is to direct, or block, RNA polymerase into transcribing certain genes based on a combination of signals received from the cell's environment.

One common mechanism for transcriptional activators to recruit RNA polymerase to the promoter region of the gene to be transcribed, or for repressors to block RNA polymerase more efficiently, is by DNA looping. By being tethered at one end to a secondary binding site, the local concentration of an activator or repressor to its primary binding site is increased tremendously. In the case of transcriptional activation of *glnA* (glutamine synthetase) during nitrogen starvation, the activator NtrC, bound at strong sites located at 110 and 140 bp upstream of the transcription start site for *glnA*, has to bind to the σ^{54} subunit of the RNA polymerase so the latter can be converted to an open complex, thereby activating transcription of the gene (Sasse-Dwight & Gralla, 1988; Echols, 1990; Weiss, Batut, Klose, Keener, & Kustu, 1991). In the arabinose operon, which contains divergent promoters p_{BAD} and p_C, positive and negative regulation of the *araBAD* genes is controlled by the AraC protein. In the absence of arabinose, AraC interacts with both the O_2 and I_2 operators, forming a 210-bp DNA loop that effectively blocks RNA polymerase access to the p_{BAD} and p_C promoters. In the presence of arabinose, AraC binds instead to the adjacent I_1 and I_2 sites, releasing the loop and activating the expression of *araBAD* genes (Matthews, 1992; Schleif, 2000). In the lactose operon, auxillary operators both upstream and downstream of the primary operator ensure high effective concentration of the lactose repressor protein, LacI, to the primary operator for efficient repression of the *lacZYA* genes in the absence of lactose (Mossing &

Record, 1986; Krämer et al., 1987; Oehler, Eismann, Krämer, & Müller-Hill, 1990). The DNA loop bridged by LacI is one of the simplest looping systems and has been the paradigm for several studies on looping (see Figure 1.1).

1.2 Thermodynamic models on DNA looping

DNA looping, as a regulatory mechanism for gene expression, requires bending and contorting the DNA. Thus, it is essential to consider the mechanical properties of this biological polymer in order to study DNA looping. The Worm-Like Chain (WLC) model has been used to describe semi-flexible polymers (Kratky & Porod, 1949), and has been successfully applied to DNA bending behavior. In this model, the energy it takes to bend DNA is dependent on the angles between planes of adjacent base pairs in the double helix and the number of base pairs. The winding of the DNA strands along the axis dyad, forming the double helix, as well as the base pair stacking interactions, confer torsional stiffness and bending rigidity to the dsDNA. This led to the definition of the intrinsic inflexibility of DNA called the persistence length. The physical implications of the WLC model on DNA looping are (1) DNA behaves as a rigid rod at contour lengths shorter than its persistence length, thus requiring considerable energy to bend and form the DNA loop; and (2) at contour lengths greater than the persistence length, DNA behaves like a random coil, in which entropy dominates over bending energy in loop formation.

T4 ligase-mediated DNA cyclization experiments has shown that the ring closure probabilities, which correlate to DNA bending energies, depend on the effective

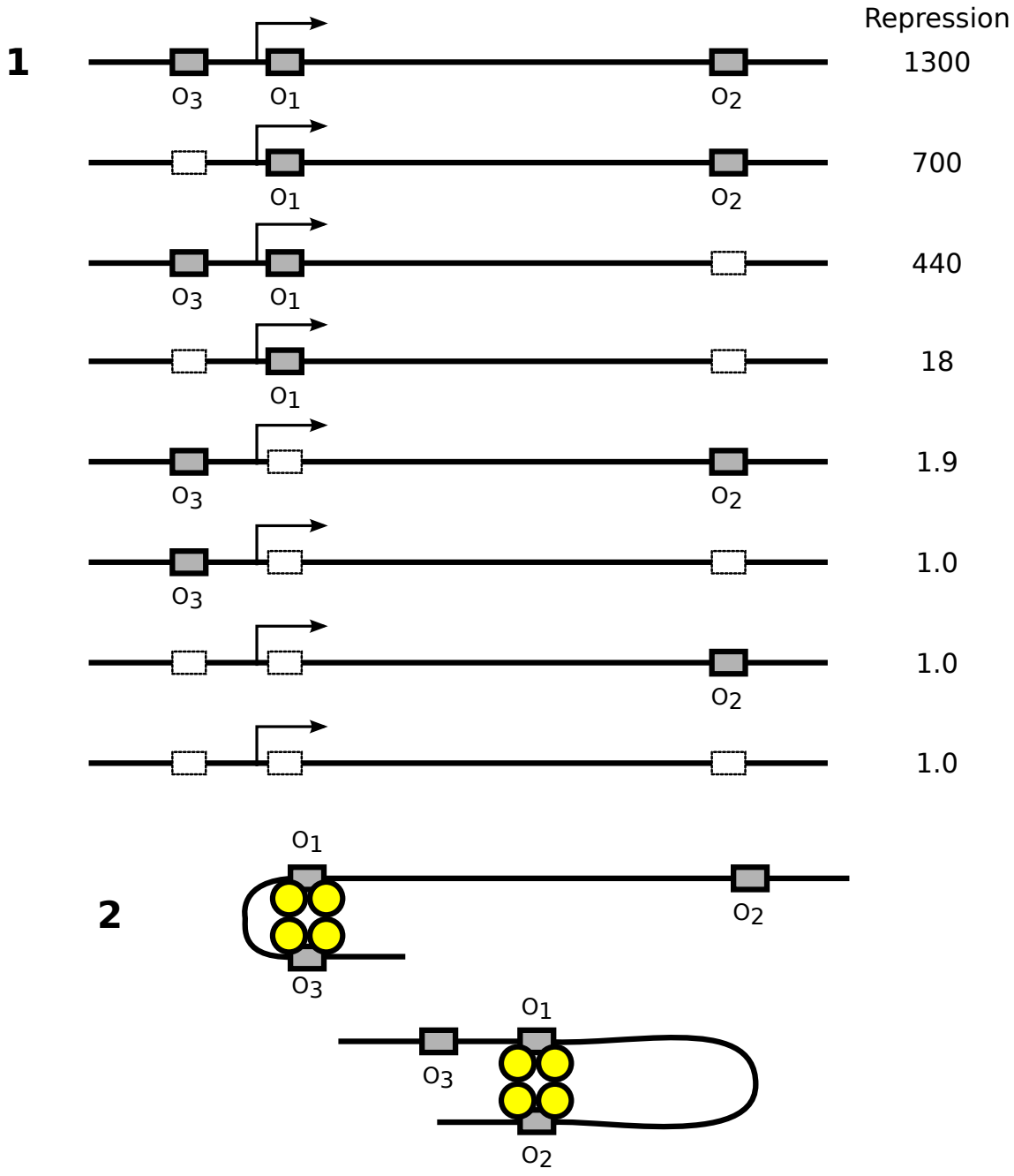


Figure 1.1: Negative regulation of the *lac* operon using LacI protein, from Oehler et al. (1990). (1) Presence of auxillary operators O_2 and O_3 enhances the repression of the *lacZYA* genes in the absence of lactose. (2) By tethering LacI (shown as four yellow circles) to either O_2 or O_3, the local concentration of LacI to the primary operator is increased by looping the intervening DNA.

equilibrium concentration of the ligatable ends in the vicinity of one another (Shore, Langowski, & Baldwin, 1981; Shimada & Yamakawa, 1984; Zhang & Crothers, 2003) (see Figure 1.2). Mossing and Record (1986) first demonstrated parallels between cooperative repression in the presence of an operator at various distances upstream of a promoter-constitutive operator control region, and length dependent circularization of DNA fragments using T4 ligase. The observed effects of DNA twist, shape and stiffness from these experiments initiated the development of a quantitative model that would explain the regulatory action of distant DNA sites. According to this model, the effective local concentration, known as the j-factor, of looping components bound to different sites in the DNA sequence is controlled by the length of the intervening DNA and its inherent bending and torsional stiffness. Measurement of the looping efficiency as a function of DNA length showed that the effective concentration increases as the distance between looping partners increases, reaching a maximum value at around 400 bp, then gradually decreases as the distance increases further. Furthermore, periodic fluctuations in the j-factor are superimposed on the overall curve, where the period of the oscillations is equivalent to one helical repeat. This is due to the specific spatial relationship of the looping partners upon binding. This model had been consistent in interpreting early experimental results on looping efficiency.

Kramer and co-workers used electron microscopy, non-denaturing polyacrylamide electrophoresis and DNAse I footprinting experiments to show that the stability of the LacI-DNA loops *in vitro* depends on the distance between the operators, the phasing of the two *lac* operators and the concentrations of repressor and DNA

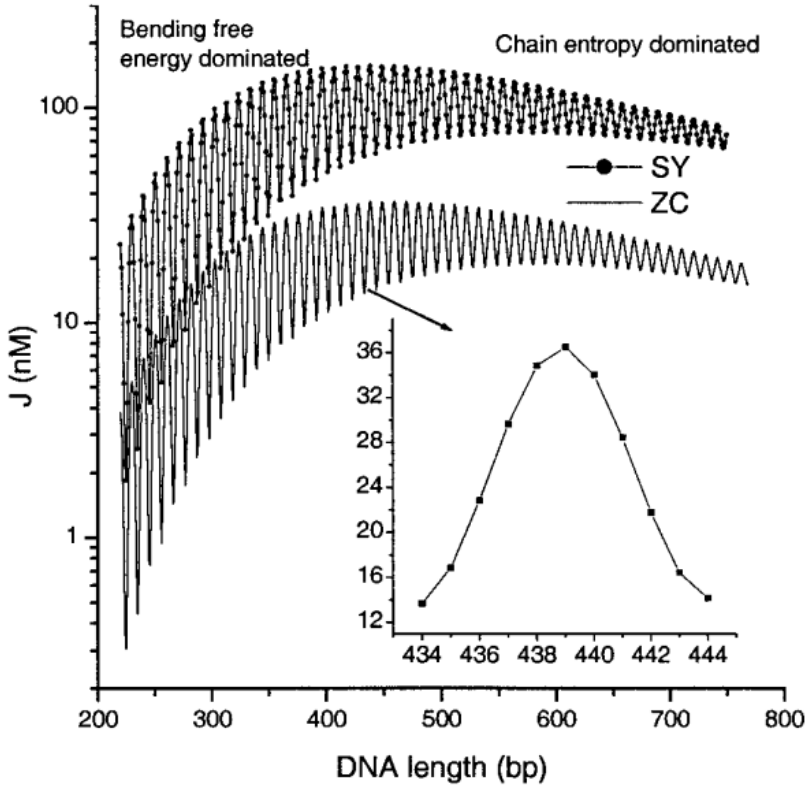


Figure 1.2: Themodynamic model for DNA cyclization, from Zhang and Crothers (2003). Calculations done by Zhang and Crothers (2003) (ZC) and Shimada and Yamakawa (1984) (SY) on homogeneous DNA were compared. Discrepancies in the calculated values of j-factors aside, this model illustrates the physical forces involved in the formation of circular DNA based on DNA length. At short DNA lengths, DNA flexibility plays a major role in forming circular DNA, whereas at longer lengths, random motion of the ligatable ends determines cyclization. Oscillations in the j-factors correspond to one helical repeat, which are due to varying orientations of the ligatable ends to one another.

(Krämer et al., 1987). Bellomy and colleagues developed a DNA looping assay, using plasmid constructs containing a wild-type *lac* operator located at various distances upstream of an operator-constitutive lac control element, to determine the looping efficiency *in vivo* (Bellomy, Mossing, & Record, 1988). Maher and colleagues developed a similar repression assay that probes properties of *lac* repression loops *in*

vivo, using single-copy reporter constructs containing varied spacings of the *lac* operators (Becker et al., 2005). Figure 1.3 shows the correlation between cooperative repression and the *lac* operator spacings. However, derived looping parameters from the *in vivo* studies suggest that *in vivo* formation of DNA loops seems to be less sensitive to the physical properties of DNA.

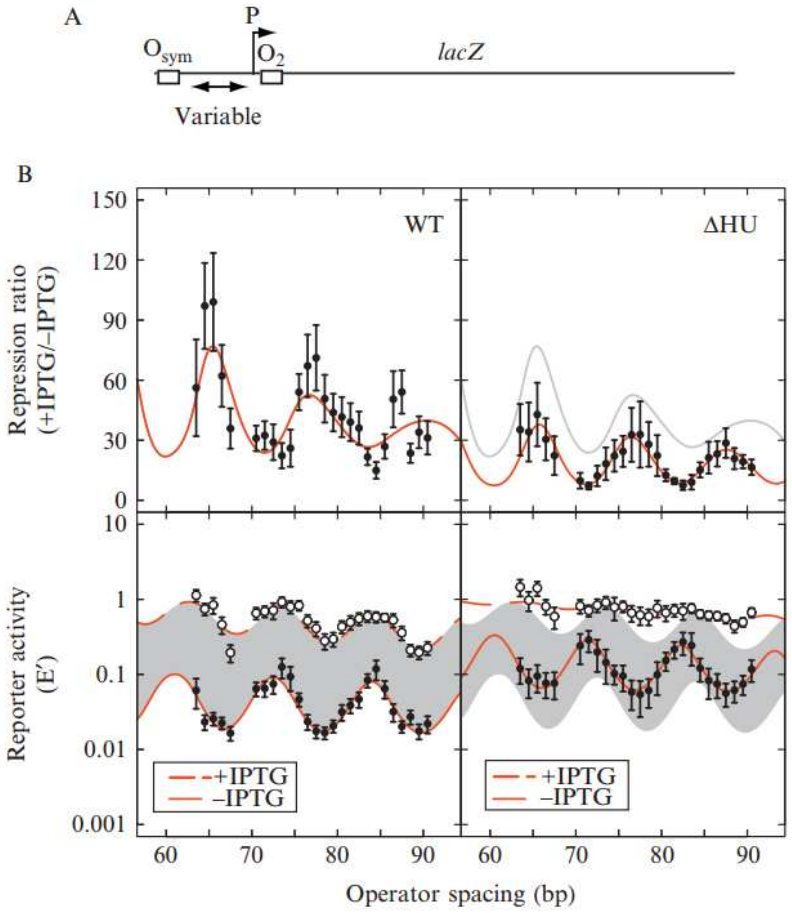


Figure 1.3: Cooperative repression via DNA looping, from Peters et al. (2011). (A) Single-copy reporter constructs used to measure repression of *lacZ* by occupancy of O_2, as a function of increased local concentration of lacI due to collisions with lacI tethered at the O_{sym} operator. (B) Experimental data fitted to a thermodynamic model for repression ratio (top) and DNA looping-dependent reporter activity, E' (bottom), from wild-type (WT) and nucleoid heat unstable HU-deficient (ΔHU) *E. coli* cells. The oscillations in all data fits demonstrate *lacZ* repression by DNA looping.

There are at least three possible explanations regarding the discrepancy in the looping efficiency between the *in vitro* and *in vivo* experiments. In the case of the NtrC-Eσ^{54} interaction, the activator is a hexamer, and so at least six possible loops can occur upon binding to DNA (Lilja, Jenssen, & Kahn, 2004). Studies probing the various loop conformations mediated by LacI suggest that the flexibility of the looping protein has a significant contribution to the apparent DNA flexibility (Vanzi, Broggio, Sacconi, & Pavone, 2006; Rutkauskas, Zhan, Matthews, Pavone, & Vanzi, 2009; Haeusler et al., 2012; Goodson, Wang, Haeusler, Kahn, & English, 2013). The bacterial assay developed by Maher and colleagues verified how architectural proteins like bacterial HU and mammalian HMGB1 affected the apparent flexibility of DNA *in vivo*. Moreover, phasing of the operators affects the looping efficiency, although a striking observation was noted in that inter-operator distances lower than the persistence length have little or no apparent effect on looping, even on systems lacking the HU protein (Becker et al., 2005; Becker, Kahn, & Maher, 2007). Taken all of these possibilities together, one can conclude that the inherent inflexibility of DNA, as described by classical elastic models and demonstrated in *in vitro* studies, is overcome in *in vivo* loop formation by interactions with non-specific DNA bending proteins, as well as the flexibility of the looping protein. DNA flexibility, in this case, plays a passive role in the formation of DNA loops.

However, spontaneous *in vitro* cyclization of 89-105 bp DNA fragments gave way to alternative models for DNA bending at DNA lengths shorter than its *in vitro* persistence length (Cloutier & Widom, 2004, 2005). One model proposed that localized melted regions within the dsDNA account for the enhanced circularization of

very short DNA fragments. Another model suggested that formation of local kinks in the intervening DNA also accounts for DNA hyperflexibility in the formation of DNA minicircles (Wiggins, Phillips, & Nelson, 2005). Both of these models indicate that DNA flexibility has a more significant role in stabilizing small *in vivo* loops. This view has been openly challenged by different research groups. Du, Smith, Shiffeldrim, Vologodskaia, and Vologodskii (2005) demonstrated that incorporation of kinks in the intervening DNA produced the enhanced cyclization efficiencies in 100-bp DNA minicircles, but failed to find a model that accounted for the periodic fluctuations in the j-factor values. Gowetski, Kodis, and Kahn (2013) designed and characterized small and relatively inflexible looping peptides to directly probe inherent DNA flexibility in DNA loop formation. Ring closure experiments on DNA fragments bound at both ends of the looping protein revealed the minimum separation of DNA binding sites to be much longer than the *in vitro* DNA persistence length. Extending these experiments on DNA fragments with varying binding site separations showed periodic fluctuations in topoisomer distributions that are consistent with the classic model for DNA bending.

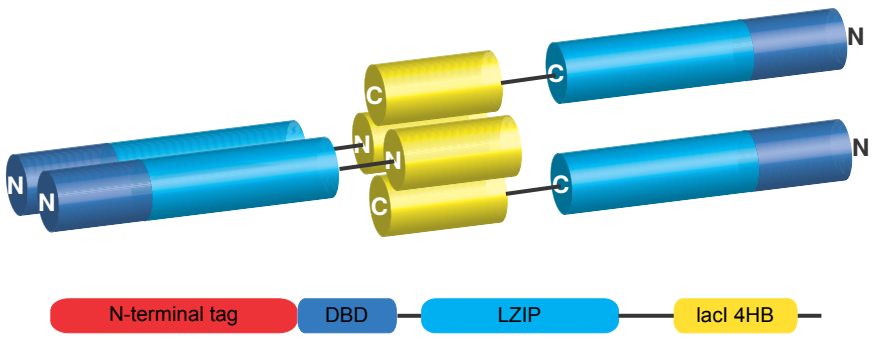
These two competing theories that explain the contribution of inherent DNA flexibility in *in vivo* loop formation is far from being resolved. Thus, we set out to design molecular tools with the intent of determining whether protein flexibility or DNA flexibility drives the formation and stability of DNA loops *in vivo*.

1.3 Approach: Removing the protein flexibility allows us to probe DNA flexibility

Most of the highly efficient looping proteins in nature are relatively large with flexible regions that allow them to have different conformations. Thus, we needed to design a new set of looping proteins that are small and relatively rigid. The looping protein has to be relatively small and inflexible such that characterization of the looping parameters would be attributed to the flexibility of the intervening DNA. Among the different protein secondary structures that can be relatively inflexible, an alpha-helical coiled-coil motif seems most promising.

Figures 1.4 and 1.5 show the schematic design for two types of short and relatively rigid proteins. The first design (LZT proteins) fuses together the dimeric DNA-binding and leucine zipper regions (LZIP) of the yeast transcription factor GCN4, and the C-terminal tetramerization domain (4HB) of the lactose repressor protein. The other design (LZD proteins) places the DNA-binding domains of GCN4 on both the N-terminal and C-terminal ends of the leucine zipper motif.

Four variants were designed for the LZT peptides by varying the sequences of the linker region between the LZIP and the lacI four helix bundles. Two of these variants (4har and 4hee) have linker sequences from the extension of the 4HB domain. The other two variants (lzar and lzee) have linker sequences from the extension of the LZIP region. The hope is that one of these variants can fold into a stable tetramer upon binding to DNA. Initial cloning attempts of these genes into



N-terminal tag: N-term--MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSDPAAL-
4har:
-KRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVARLARQVRALADSLMQLARQVSRLADSL--C-term
4hee:
-KRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVEELARQVRALADSLMQLARQVSRLADSL--C-term
lzar:
-KRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVARLKKLVRALADSLMQLARQVSRLESGQ--C-term
lzee:
-KRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVEELKKLVRALADSLMQLARQVSRLESGQ--C-term

Figure 1.4: LZT protein design. top: schematic design for the LZT proteins, indicating the DNA binding domain (dark blue rods), dimerization domain (light blue rods), the linker sequence (black lines), and the tetramerization domain (yellow rods). middle: components of the LZT protein monomer, cloned with the N-terminal affinity tag. The linker sequences between LZIP and lacI 4HB were determined by extending the 4HB domain (in 4har and 4hee), or extending the LZIP region (in lzar and lzee). bottom: amino acid sequences of the four variants of the LZT proteins, cloned as 6x His fusion proteins.

pGEX2T plasmid resulted in several mutations in the gene inserts. The genes were eventually fixed and were successfully cloned into pRSETA.

On the other hand, three variants were cloned for the LZD peptides, by extending the LZIP sequence by one or two heptad repeats. As illustrated in Figure 1.6, changing the linker sequence length rotates the orientation of one of the binding sites in a left-handed twist, giving rise to different crossover angles between DNA binding sites.

Two of the LZD peptides (LZD73 and LZD87) have been previously demonstrated to stabilize DNA loops *in vitro* (Gowetski et al., 2013). However, a closer

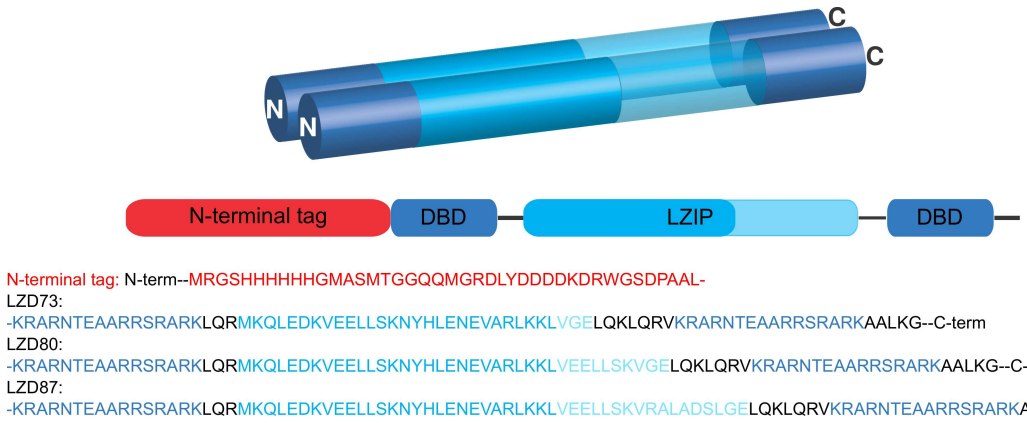


Figure 1.5: LZD protein design, from Gowetski et al. (2013). top: schematic design for the LZD proteins, indicating the DNA binding domains (dark blue rods) and dimerization domain (opaque and translucent light blue rods). middle: components of the LZD protein monomer, cloned with the N-terminal affinity tag. The linker sequence between LZIP and C-terminal DBD was determined in Hollenbeck et al. (2001). bottom: amino acid sequences of the three variants of the LZD proteins, cloned as 6x His fusion proteins.

look on the topoisomer distributions in ring closure experiments revealed similar peak locations for the different topoisomers for both peptides, suggesting similar torsional flexibilities for peptides that were designed to have different crossover angles of the DNA binding sites. By utilizing molecular tools that probe *in vivo* DNA looping by these peptides, we might be able to differentiate the *in vivo* looping abilities among the LZD variants.

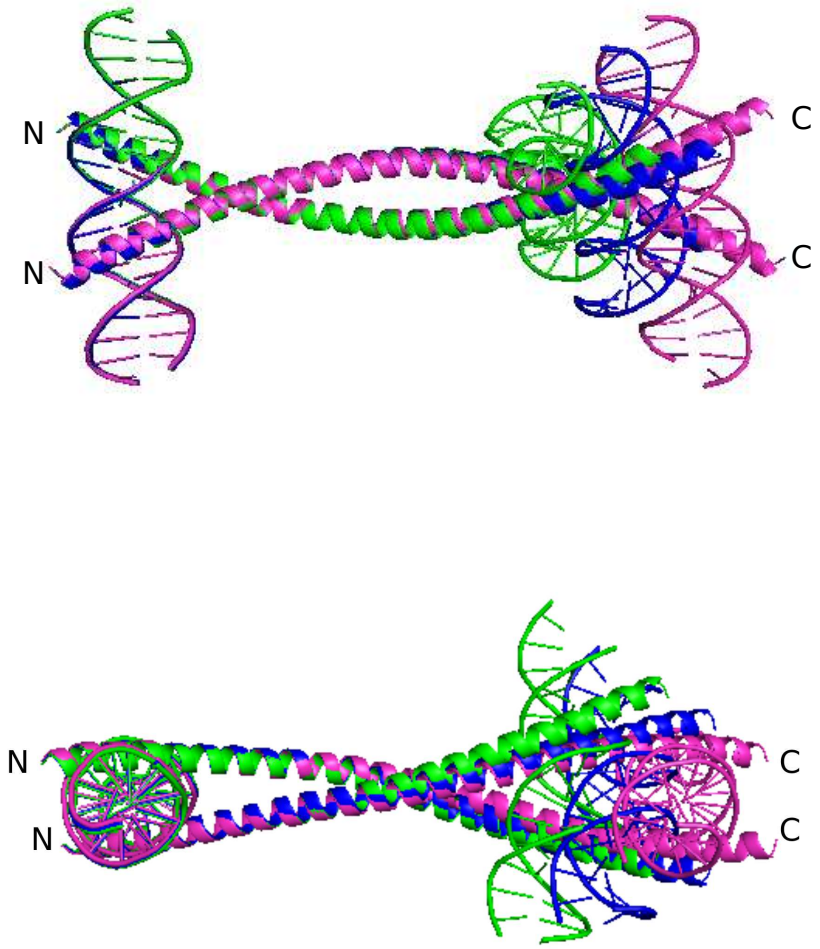


Figure 1.6: Graphical renderings of LZD73 (green), LZD80 (blue) and LZD87 (magenta) bound to 20-bp DNA containing CREB or INV2 site. The image illustrates the left-handed twisting of the coiled-coil domain. As the linker sequence is extended, the orientation of the binding sites rotates in a left-handed direction. Composite images were generated using PyMOL, with PDB file 2DGC as template.

Chapter 2: Initial expression and *in vitro* characterization of the LZT and LZD proteins

2.1 Objectives

Initial attempts to clone the LZT proteins in *E. coli* resulted in various gene sequence deletions, which indicated that baseline expression levels of these proteins are toxic to the bacterial host. The mutated gene sequences were eventually fixed using PCR and cloned into T7 promoter-controlled vectors. However, subsequent protein expression studies showed that three of the proteins (4har, 4hee and lzar) form insoluble aggregates. This makes protein characterization by *in vitro* experiments rather cumbersome although not impossible. Thus, we present two alternative protein expression strategies to produce soluble recombinant proteins. The first strategy is to express the putative proteins in insect cells using recombinant baculoviruses. The second is to attach a GST affinity tag to the recombinant proteins, and express them in bacterial cells. Furthermore, we present initial characterization results for two of the LZT proteins in their ability to loop DNA *in vitro*. The main goals addressed in this chapter are to (1) express and isolate soluble LZT proteins, and (2) to test their ability to loop DNA *in vitro*.

2.2 Rationale

Our first strategy in producing soluble artificial DNA-binding proteins is to express them in eukaryotes. Compartmentalization of different cellular processes in eukaryotes would prevent our DNA binding proteins from interacting with the chromosomal DNA. Among the various eukaryotic expression systems available, we decided to use recombinant baculoviruses to generate our artificial proteins. The relative ease in generating recombinant baculoviral vectors, along with the eukaryotic post-translational processes of the host make this system ideal in producing our proteins.

Baculoviruses are rod-shaped viruses that primarily infect insect larvae. Two of the most common viruses used in recombinant gene expression are *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) and *Bombyx mori* nuclear polyhedrosis virus (BmNPV). The life cycle of the wild-type baculovirus is shown in Figure 2.1.

In recombinant gene expression using baculoviruses, the lytic cycle is utilized to generate high-titer viral stock solutions to be used for protein expression studies. To generate the desired protein, the recombinant gene is placed downstream of the polyhedrin promoter so that the gene product is expressed instead of the polyhedrin protein. Furthermore, the polyhedrin gene product in the wild-type baculovirus makes up around 50% of the total viral protein. Thus we would expect to have reasonable yield of the recombinant protein.

Glutathione-S-transferases, on the other hand, are a family of eukaryotic and

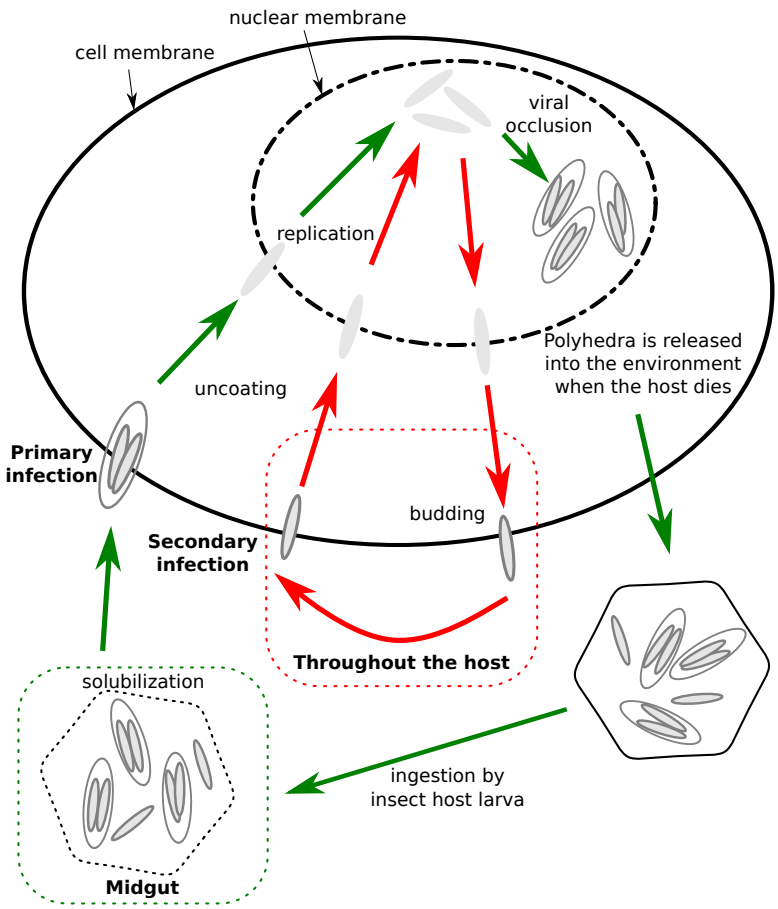


Figure 2.1: Baculovirus life cycle in insect cells from Murphy et al. (2004). Viral infection of the insect larva generates two types of progeny, which are developed independently throughout the life cycle. During the lytic cycle (shown in red arrows), the host cell produces non-occluded or budded viruses. The budded virus is coated with a lipid bilayer membrane acquired from the host cell membrane. This promotes horizontal transmission of the virus to other types of cells in the larva. On the other hand, occluded viruses are generated during the occluded cycle (shown in green arrows). The mature virions are packaged together inside an envelope acquired from the nuclear membrane of the host, forming the occluded virus. Several occluded viruses are then embedded within a polyhedrin crystal matrix. When the larva dies, the polyhedra are released into the environment where they can then be ingested by another insect host, promoting the vertical transmission of the virus.

prokaryotic metabolic enzymes whose known purpose is to remove xenobiotics and products of oxidative stress by conjugating them with the reduced form of glutathione (GSH) (Hayes, Flanagan, & Jowsey, 2005). In recombinant gene tech-

nology, *Schistosoma japonicum* GSTs are cloned into the recombinant protein as an N-terminal affinity tag (Smith & Johnson, 1988). This strategy has a two-fold advantage: (1) its strong affinity to glutathione makes it relatively easy to isolate and purify the protein of interest from bacterial lysates, and (2) dimerization of the affinity tag (Parker, Lo Bello, & Federici, 1990; Ji, Zhang, Armstrong, & Gilliland, 1992; Maru, Afar, Witte, & Shibuya, 1996) would aid in the solubilization of some recombinant proteins deemed toxic to the host cells, as well as trigger correct folding of the LZIP motif.

Finally, we used Electrophoretic Mobility Shift Assay (EMSA) to see whether our artificial proteins can actually loop DNA. EMSA, also called gel shift assay, is based on the premise that DNA bound to a protein will migrate slower in a polyacrylamide gel, when an electric field is applied to it. To illustrate the ability to loop DNA, a DNA-looping protein should first have the ability to bind two separate DNA strands, forming a structure called "sandwich complex." Binding reaction conditions are optimized such that the formation of a protein-DNA sandwich complex is highly dependent on the correct stoichiometry between the protein and its DNA binding site.

2.3 Materials and Methods

Reagents used for microbiology work were purchased from Fisher Scientific, Sigma, and USB Corporation. The oligonucleotides used for cloning were synthesized by Integrated DNA Technologies (IDT). All molecular biology reagents and

enzymes were purchased from New England Biolabs (NEB). The pFastBac1 plasmid (Invitrogen) is a kind gift from the Bentley lab, while pGEX6P1 was purchased from Amersham Biosciences. PCR was done on either Eppendorf Mastercycler or MJ Research PTC-200 thermal cycler. All protein purifications were done on the ÅKTA FPLC instrument (Amersham Biosciences), attached with either a 1-mL prepacked HisTrap HP (GE Healthcare) column previously charged with 0.2 M cobalt acetate, or a 1-mL prepacked GSTrap FF (GE Healthcare) column.

Native PAGE gels for EMSA were run on a Hoefer SE600 Dual Cooled vertical electrophoresis unit. Phosphorimages were taken using the Storm 860 Phosphorimager, and processed using ImageJ software.

2.3.1 Subcloning into pFastBac1 and pGEX6P1 plasmids

The DNA sequences for the tetrameric DNA binding proteins, which were previously cloned into pRSETA (see Appendix 2.1, pages 146 and 153 for gene sequences of pRSETA and LZT gene inserts), were amplified by PCR using the following primers: 5'-TATAGGGAGCTCACAACGGTTTCC-3' and 5'-TGCTAGT TATTGCTCAGCGGTGG-3'. The resulting PCR product was digested with either *Sac*I and *Hin*dIII, or with *Bam*HI and *Eco*RI, and purified using non-denaturing PAGE. In parallel, pFastBac1 was digested with *Sac*I and *Hin*dIII, while pGEX6P1 was digested with *Bam*HI and *Eco*RI. The digestion products of the parent plasmids were treated with phosphatase to prevent religation. The recombinant plasmids were subsequently assembled by ligating together the digestion products of the gene insert

to the parent plasmids.

The ligation products were then transformed into *E. coli* XL1-Blue electrocompetent cells for recombinant plasmid propagation using standard electroporation protocols. Transformants were plated onto LB agar plates containing 100 μg/mL ampicillin. Colonies were then selected for liquid culture growth in LB amp media. Plasmid preps were done using QIAprep Spin miniprep kit, according to manufacturer's instructions. Correct clones were identified by *Ban*I (for pFastBac1 clones), or *Pvu*II (for pGEX6P1 clones) digestion, and by DNA sequencing.

2.3.2 Protein expression using baculoviruses

To generate the baculoviruses expressing our DNA binding proteins, the recombinant pFastBac1 plasmids (see Appendix 2.1, page 147) were transformed into DH10Bac *E. coli* cells according to manufacturer's transformation protocols, and plated onto LB agar plates containing 50 μg/mL kanamycin, 7 μg/mL gentamicin, 10 μg/mL tetracycline, 100 μg/mL Bluo-gal and 40 μg/mL IPTG. White colonies were selected for liquid culture growth in LB kan gen tet media, and the baculovirus shuttle vector (bacmid) DNA was harvested using standard DNA prep protocols. In brief, cells were resuspended in Solution I (15 mM Tris-HCl pH 8, 10 mM EDTA, 100 μg/mL RNase A), and lysed in Solution II (0.2 N NaOH, 1% SDS). Host proteins and chromosomal DNA were precipitated out using 3M KOAc pH 5.5, and the supernatant was treated with isopropanol and 70% ethanol to isolate the bacmid DNA. Finally, the bacmid DNA was resuspended in TE (10 mM Tris-HCl pH 8, 1

mM EDTA) buffer. Recombinant bacmids were then analyzed by PCR using M13 Forward (-41) [5'-CGCCAGGGTTTTCCCAGTCACGAC-3'] and M13 Reverse (-27) [5'-CAGGAAACAGCTATGAC-3'] primers.

The recombinant bacmids were subsequently transfected into *Sf9* insect cells to generate P1 and P2 baculoviral stock solutions. The protocol for generating P1 stock solutions was as follows: *Sf9* cells were allowed to settle at the bottom of a 6-well culture plate (9 x 10^5 cells per well, containing 2 mL Sf900-II SFM + 50 μg/mL streptomycin) at 27 °C prior to baculoviral transfection. To prepare the baculovirus, 1 μg of purified bacmid was coated with Cellfectin® reagent before mixing into the prepared cell hosts. Transfection was done at 27 °C for 5 hours, then the cells were supplemented with growth media (Sf900-II SFM + 50 μg/mL streptomycin), and incubated at 27 °C for 72 hours. The P2 baculoviral stock solutions were prepared by inoculating 2 x 10^6 cells with P1 stock at 0.1 pfu/cell and incubated at 27 °C for four days. The number of infected cells was determined by a viral titer, and the P2 viral stock (cells and growth media) was analyzed by SDS-PAGE and Western blot to determine whether the expressed protein is being secreted out of the cell hosts.

For large-scale expression of the tetrameric DNA binding proteins, at least 30 mL of *Sf9* cells were transfected with the P2 viral stock at MOI = 1. The transfected cells were incubated at 27 °C for 72 hours with aeration before cell harvest. The cells were resuspended in Lysis Buffer (25 mM KOAc pH 5.5, 500 mM NaCl, 20 mM imidazole), and lysed by repeated freeze-thaw cycles and several passes through a 25G needle. The lysate was cleared by centrifugation (10000 x g, 4 °C, 15 minutes; supposedly at 15000 x g, 4 °C, 1 hour) and filtered through a 0.2-μm syringe filter

prior to purification by affinity chromatography.

2.3.3 Protein expression in BL21 *E. coli* cells

Recombinant pGEX6P1 plasmids (see Appendix 2.1, pages 150 and 153) were transformed into BL21 electrocompetent cells, according to standard electroporation protocols, and plated onto LB agar plates containing 20 mM glucose and 100 μg/mL ampicillin. Colonies were selected for liquid culture growth in 10 mL 2x YT + 100 μg/mL ampicillin, incubated at 37 °C with ample aeration for at least 12 hours. The liquid cultures were scaled up to 2 L 2x YTA for 3 hours at 37 °C, and protein expression was induced for 20 hours at 37 °C with 0.1 mM IPTG. The cells were then harvested by centrifugation (7700 x g, 4 °C, 10 minutes), resuspended in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3), and lysed by three passes through the French press (10000-12000 psi). The soluble and insoluble lysates from each pass were analyzed by SDS-PAGE to determine optimal lysis conditions. After lysis, Triton X-100 was added to 1% final concentration and the lysate was cleared by centrifugation (>10000 x g, 4 °C, 20 minutes) and filtered through a 0.2-μm syringe filter prior to purification by affinity chromatography.

2.3.4 Protein expression in BL21-AI *E. coli* cells

Recombinant pRSETA plasmids were transformed into BL21-AI electrocompetent cells, according to standard electroporation protocols, and plated onto LB agar plates containing 100 μg/mL ampicillin. Colonies were selected for liquid cul-

ture growth in 10 mL LB + 20 mM glucose and 100 μg/mL amp and incubated for at least 12 hours at 37 °C. The starter cultures were then used to inoculate 500 mL LB amp, and incubated at 37 °C with aeration until the OD_{600} reached 0.4 – 0.6 (1 – 3 hours). Protein expression was induced with 0.02% arabinose at 37 °C for 18 hours with aeration. The cells were harvested by centrifugation, resuspended in Equilibration Buffer (10 mM MES-NaOH pH 6, 500 mM NaCl, 20 mM imidazole), and lysed by three passes through the French press. The crude lysate was cleared by centrifugation (14000 x g, 4 °C, 1 hour) and filtered through a 0.2-μm syringe filter prior to purification by affinity chromatography.

2.3.5 Protein isolation, purification and workup

Two buffer systems were developed in the course of optimizing pH and salt conditions for isolating the DNA binding proteins containing the His affinity tag. The following buffers were used for the isolation of the His-tagged proteins harvested from *Sf9* cells:

- Equilibration – 25 mM KOAc pH 5.5, 500 mM NaCl, 20 mM imidazole
- Elution – 25 mM KOAc pH 5.5, 500 mM NaCl, 500 mM imidazole

Isolation of the His-tagged proteins harvested from BL21-AI cells was done using the following buffers:

- Equilibration – 10 mM MES-NaOH pH 6, 500 mM NaCl, 20 mM imidazole
- Elution – 10 mM MES-NaOH pH 6, 500 mM NaCl, 600 mM imidazole

Lastly, the buffers used for isolating GST-tagged proteins are the following:

- Equilibration – 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3 (PBS)
- Elution – 100 mM Tris-HCl pH 8, 50 mM reduced glutathione

After sample loading, a 5-CV wash with Equilibration buffer was passed through the column. The protein was then eluted out of the column in 1-mL aliquots either by an imidazole gradient (20 – 400 mM or 20 – 600 mM), or step elution using 50 mM reduced glutathione through 25 or 50 CVs. The eluates were analyzed by SDS-PAGE, in 18% (32:1) or 16% (40:1) polyacrylamide gels, using a Tris-glycine-SDS or Tris-tricine-SDS buffering system ran at 25 W for 50-60 minutes, to check for purity. The presence of tricine in the running buffer allows for better separation of low MW proteins (Schägger & von Jagow, 1987). Eluates containing pure His-tagged protein were concentrated and buffer-exchanged with Storage Buffer (20 mM HEPES-NaOH pH 7, 150 mM NaCl, 10% glycerol) using Amicon YM-10 membrane filters, and stored at -80 °C.

The affinity tags from GST-tagged proteins were cleaved off using the following protocol: the pooled samples from the FPLC run were first buffer exchanged with PreScission Cleavage Buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) through a 5-mL prepacked HiTrap Desalting Column (GE Healthcare). Eluents containing the protein were pooled and treated with PreScission protease at 4 °C for 4 hours. Finally, the cleaved GST tag and PreScission protease, which is also GST-tagged, were isolated from the mixture by FPLC through a 1-mL prepacked

GSTrap FF column, as the untagged protein is collected in the flowthrough.

2.3.6 Purification of His-tagged fusion proteins from inclusion bodies

The insoluble pellet after the French press cell lysis step was resuspended in Solubilization Buffer (10 mM MES-NaOH pH 6, 500 mM NaCl, 6 M Gua-HCl, 5 mM imidazole), and incubated at 4 °C for at least 30 minutes with occasional mixing until a homogeneous mixture is obtained. The mixture is then cleared by centrifugation (14000 x g, 4 °C, 1 hour), and the clarified lysate was loaded manually (1 drop/sec) into the 1-mL prepacked HisTrap HP column charged with 0.2 M cobalt acetate. The column was then washed manually with 10 CV of Solubilization Buffer (1 drop/sec) and attached into the FPLC instrument.

The buffers used for refolding and isolating the His-tagged proteins were as follows:

- Denaturing – 10 mM MES pH 6, 500 mM NaCl, 8 M urea, 5 mM imidazole
- Equilibration – 10 mM MES pH 6, 500 mM NaCl, 20 mM imidazole
- Elution – 10 mM MES pH 6, 500 mM NaCl, 600 mM imidazole

A 10 CV wash with Denaturing Buffer was done to remove the guanidine from the column, then the protein is refolded in-column by a linear gradient of decreasing urea concentration (6 – 0 M). The protein was then eluted out using an imidazole gradient (20 – 600 mM). The eluates were analyzed by SDS-Tricine PAGE, and fractions containing the His-tagged protein were pooled, concentrated

and buffer-exchanged with Storage Buffer using Amicon YM-10 membrane filters. The concentrated samples were then stored at -80 °C.

2.3.7 Electrophoretic Mobility Shift Assay (EMSA)

First, a 26-mer ssDNA containing the palindromic GCN4 site CREB was end-labeled with γ-^{32}P ATP in a 15-μL reaction mixture that contains the following: 5 μM CREB DNA (5'-CAGTGGAGATGACGTCATCTCGTGCC-3'), 1x T4 PNK Buffer, 100 μmol γ-^{32}P ATP, 10 U T4 PNK. The reaction was done at 37 °C for 1 hour. Afterwards, the reaction was diluted to 75 μL with STE (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM NaCl), and unincorporated label was removed by passing the sample through a P-30 MicroBioSpin column (Bio-Rad). The complementary strand was then annealed to the labeled DNA, and the dsDNA mixture was incubated in a heat block preheated to 100 °C that is gradually cooled to $\sim$ 40 °C. The labeling efficiency was then measured for a 100-fold dilution of the dsDNA by scintillation counting.

The following components of the binding reaction (10-μL total volume) were mixed together in a buffer formulation containing 20 mM Tris-HCl pH 7.5, 60 mM KCl, 5 mM $MgCl_2$, 10 mg/mL BSA, 1 mM DTT, 1 % glycerol, 4 % sucrose, 0.0025% bromophenol blue: 25 nM labeled CREB DNA, 0.5 μg poly(dI-dC), 1 μL purified protein. Protein dilutions, when needed, were done in Protein Dilution Buffer (20 mM Tris-HCl pH 7.5, 150 mM KCl, 1 mM DTT, 10 mg/mL BSA, 5 % glycerol, 0.1 % Triton X-100). The binding reaction mixture was incubated at ambient temperature

for 30 minutes, and then subsequently analyzed using native PAGE.

The following gel formulation was cast for the EMSA: 10 % (75:1) polyacrylamide gel in 1x Native PAGE Buffer (25 mM Tris-Cl, 192 mM glycine pH 8.3) supplemented with 5 % glycerol. The gel was maintained at <16 °C while running at 400 V for 1 hour. The resulting gel was then dried on a filter paper, and exposed to the phosphor screen (Molecular Dynamics) for at least 8 hours prior to image visualization and processing.

2.4 Results and Discussion

2.4.1 Protein expression and purification

Initial protein expression in *Sf9* insect cells gave good yields from all four LZT proteins. Figure 2.2 shows the results from 4har.

Next, we scaled up the protein expression for 4har to 750 mL culture, and proceeded to isolate the His-tagged protein by FPLC. The chromatogram showed a significant peak at fractions B4 to B6, and a wide and shallow peak at fractions B7 to B10, which corresponds to elution with $\sim$ 80-150 mM imidazole. However, SDS-PAGE and Western blot analyses failed to detect the His-tagged 4har.

Baculoviral expression of our recombinant proteins is expensive and time consuming, in terms of growing the host cell line to the desired cell density prior to cell transfection. Thus, we decided to do another attempt at expressing our proteins in bacteria.

Of the four LZT proteins, we were successful in subcloning 4har and lzee into

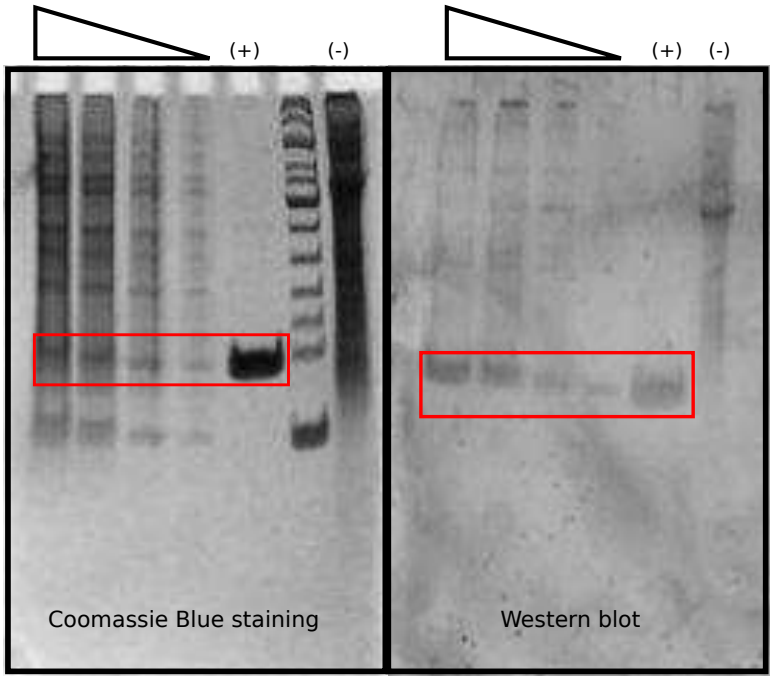


Figure 2.2: Baculoviral expression of 4har protein in *Sf9* cells. Left panel: SDS-Tricine PAGE on 20 % (32.3:1) polyacrylamide gel, stained with Coomassie Blue. Various amounts of crude lysate (lanes 1-4) illustrates the appearance of a protein band that matches the approximate size for FPLC-purified His-tagged lzee (+). Right panel: Western blot of increasing amounts of crude lysate verifies the presence of the Histidine-tagged 4har (lanes 1-4) which matches the approximate size for FPLC-purified His-tagged lzee (+). A smear on the untransfected cells (-) and high MW bands on the lysates on the Western blot are due to non-specific binding of the antibody. The protein of interest is boxed in red.

pGEX6P1. Initial protein expression of these proteins in BL21 cells showed a distinct band at around the 35 kDa protein ladder mark from the recombinant plasmids, which is roughly the size of the GST-LZT fusion proteins (∼36 kDa). As a control, protein expression from the empty vector generated a dark band that corresponds to the size of the GST tag (∼26 kDa) Furthermore, there was no significant difference in the amount of expressed proteins between 0.1 and 1 mM IPTG. Thus, we proceeded to large-scale expression using 0.1 mM IPTG and purification for both proteins.

SDS-PAGE analysis of the clarified lysates and the cell pellets for lzee from

each French press pass showed that it takes at least two passes to get decent yields of the GST-tagged protein from the soluble material. Moreover, the chromatogram from the FPLC purification of the final clarified lysate L3 revealed a distinct peak at fractions A2-A4, which was verified by SDS-PAGE to be the GST-tagged fusion protein (see Figure 2.3). Similar results were also seen in the purification of GST-4har fusion protein. Now that we have isolated GST-tagged 4har and lzee, we can then proceed to *in vitro* characterization.

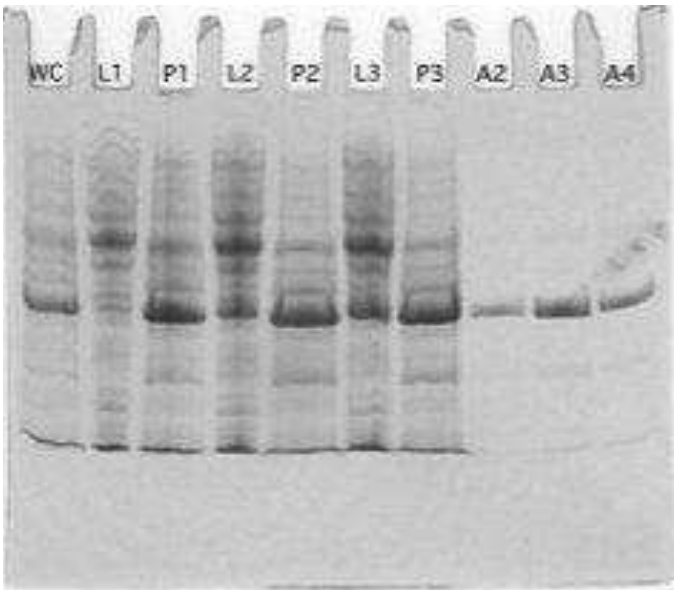


Figure 2.3: Protein expression and purification of GST-lzee in BL21 cells. It takes at least two passes through the French press to get good lysis quality, as evidenced by the increasing intensity of a protein band that lines up with the FPLC-purified GST-tagged lzee in fractions A2-A4. WC: whole cells, 20 hours post-induction with 0.1 mM IPTG at 37 °C. L1, L2, L3: clarified lysates from each pass through the French press. P1, P2, P3: cell pellets from each French press passage. A2, A3, A4: FPLC fractions purified from clarified lysate L3, containing GST-tagged lzee. All sample aliquots were taken from the same protein prep.

2.4.2 EMSA

To form DNA loops, a DNA binding protein has to bind to a single DNA strand at two different regions. However, the length of the intervening DNA and the orientation of the binding sites affect the stability of the resulting loop, and would complicate analysis of our characterization studies. Thus, we devised a simple assay that would illustrate the first requirement for DNA looping: show that our proteins can actually bind two independent DNA strands. Short oligonucleotides containing the GCN4 binding site CREB were mixed together with purified proteins. The resulting complexes were subsequently analyzed by native PAGE. The migration of a DNA-protein complex should differ from that of free DNA. Different complexes would have distinct mobilities in a polyacrylamide gel. Thus, if a protein molecule can bind to two DNA strands, forming a sandwich complex, it will have a distinct shifted band that will only show up when the stoichiometries between the protein and the DNA are optimized.

SDS-PAGE analysis of the the GST-tagged fusion proteins to be used in the gel shift assay revealed that all of the protein preps are starting to degrade into two bands (see Figure 2.4). The degradation seem to occur more rapidly when the protein samples are left at ambient temperature overnight (GST-lzee #1), as seen by the protein bands at ∼26 kDa (GST affinity tag) and ∼10 kDa (untagged LZT protein). Storage at 4 °C (GST-lzee #2 and GST-4har #1) and -20 °C (GST-lzee #3 and GST-4har #2) showed partial degradation of the ∼36 kDa band (GST-tagged fusion proteins) into two smaller protein bands. Thus, long term storage of

LZT protein preps are done at -80 °C, while the working solutions are temporarily stored at -20 °C to minimize proteolytic activities.

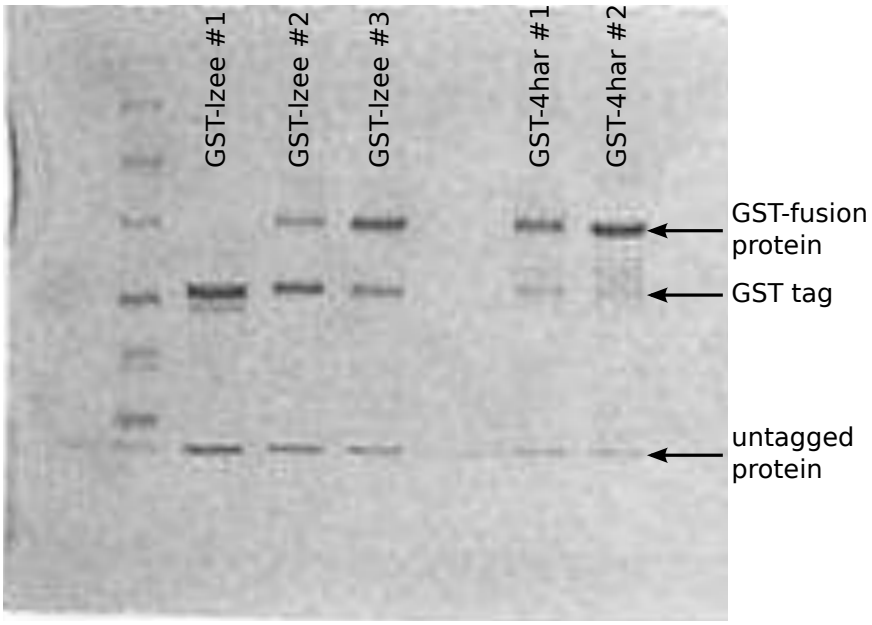


Figure 2.4: SDS-PAGE of FPLC-purified GST fusion proteins stored at different temperatures. Protein preps were exposed at ambient temperature (GST-lzee #1), 4 °C (GST-lzee #2 and GST-4har #1), and -20 °C (GST-lzee #3 and GST-4har #2) for at least 10 hours prior to analysis. Prolonged exposure to higher temperatures triggers proteolysis at the linker region between the GST tag and the protein insert.

We thought that the proteolysis would result to a partially degraded protein that can still bind to its DNA substrate. Thus, we proceeded to analyze all of our GST-fusion protein preps by EMSA.

Figure 2.5 shows the gel shift assay of the GST-tagged proteins with the 26 bp DNA oligomer containing the CREB binding site. Results showed that the two GST-lzee protein preps gave different shifted complexes at pH 7.5. GST-lzee #1 prep, which was previously exposed to ambient temperature, gave a single population of DNA-protein complexes, while GST-lzee #2 prep, stored at 4 °C, showed three

shifted bands. Furthermore, all of the populations increased linearly with a linear increase in the protein concentration. On the other hand, GST-4har showed four complexes with the topmost band having the highest intensity. Similar results were also seen from binding reactions done at pH 5.7.

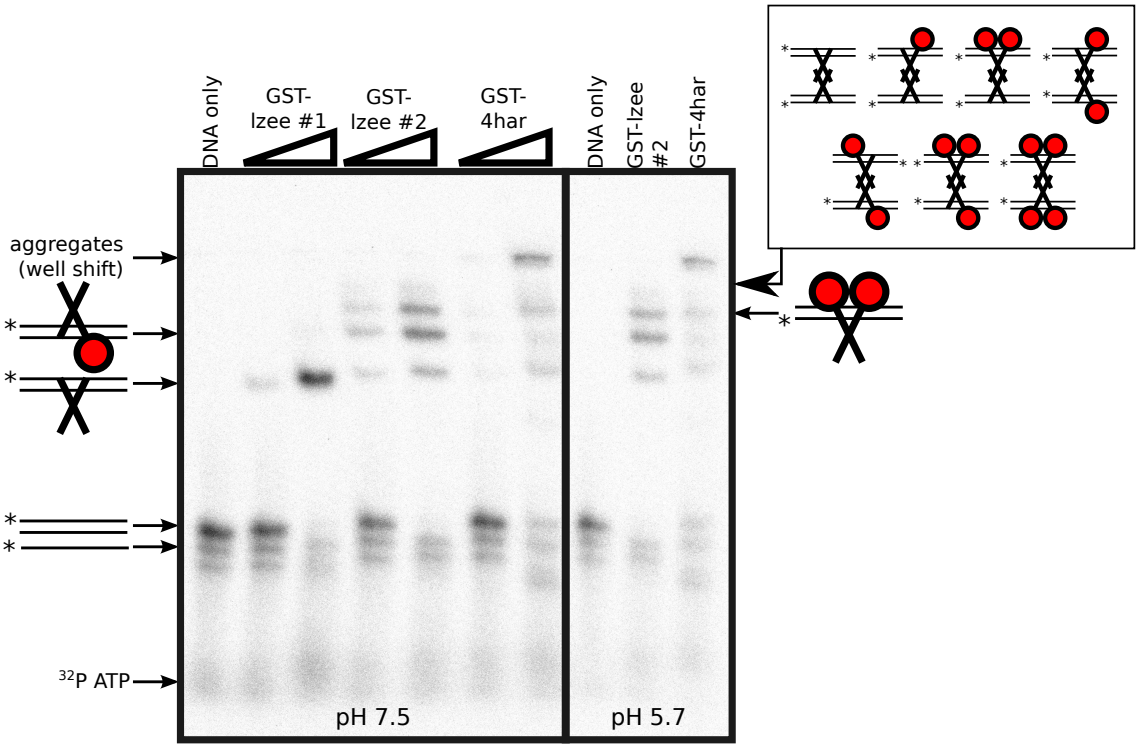


Figure 2.5: EMSA on GST-4har and GST-lzee with 26-mer CREB DNA, at different pH conditions. The GST affinity tags are shown as red circles. ^{32}P-labeled DNA in each binding reaction is 25 nM. Different shifted bands are due to the different combinations of protein dimers bound to the DNA. Any bands that are stuck on the wells are most likely from insoluble protein-DNA aggregates. There is no significant difference in the number of shifted bands between different pH conditions.

The GCN4 LZIP protein has been previously shown to bind with high affinity to the palindromic CREB site (Hill, Hope, Macke, & Struhl, 1986; König & Richmond, 1993). Given that the LZT protein preps used in the gel shift assay are essentially the extended version of GCN4 LZIP, we thought that the simplest

and most stable complex that can enter the polyacrylamide gel matrix is that of a protein dimer bound to one DNA fragment. Thus, we initially assigned the single shifted band from GST-lzee #1 to be that of the untagged lzee dimer:DNA complex. Partially degraded protein preps would then bind to DNA with different dimer combinations, giving way to three possible shifted bands. It is possible that some or all of the shifted bands are actually sandwich complexes. In order to further verify our initial assignment of the shifted bands, we cleaved off the GST tag from lzee (see Section 2.3.5, page 22 for procedure) and repeated the assay at pH 7.5.

As shown in Figure 2.6, saturating amounts of GST-4har resulted to a well shift, indicating that 4har forms aggregates along with the DNA. We speculate that the C-terminal end of the protein dimer, which does not bind DNA, forms networks with other similar complexes. On the other hand, gel shift assay results on the untagged lzee showed a single shifted band. In order to make sense of the binding behavior, we also did the assay using PAA peptide. PAA is a GCN4 derivative that has been previously shown to bind to CREB site as a dimer (Strauss-Soukup & Maher, 1997). Gel shift assay results on PAA binding to a DNA fragment containing a single CREB site showed two shifted bands of different intensities. The major band is most likely due to the PAA dimer:DNA complex. The more retarded band can be attributed to two protein dimers bound to a single DNA fragment, in which the second dimer is bound nonspecifically to the DNA. Since the untagged lzee preps showed similar band shifting, we conclude that the lzee protein can only bind to a single DNA fragment as a protein dimer.

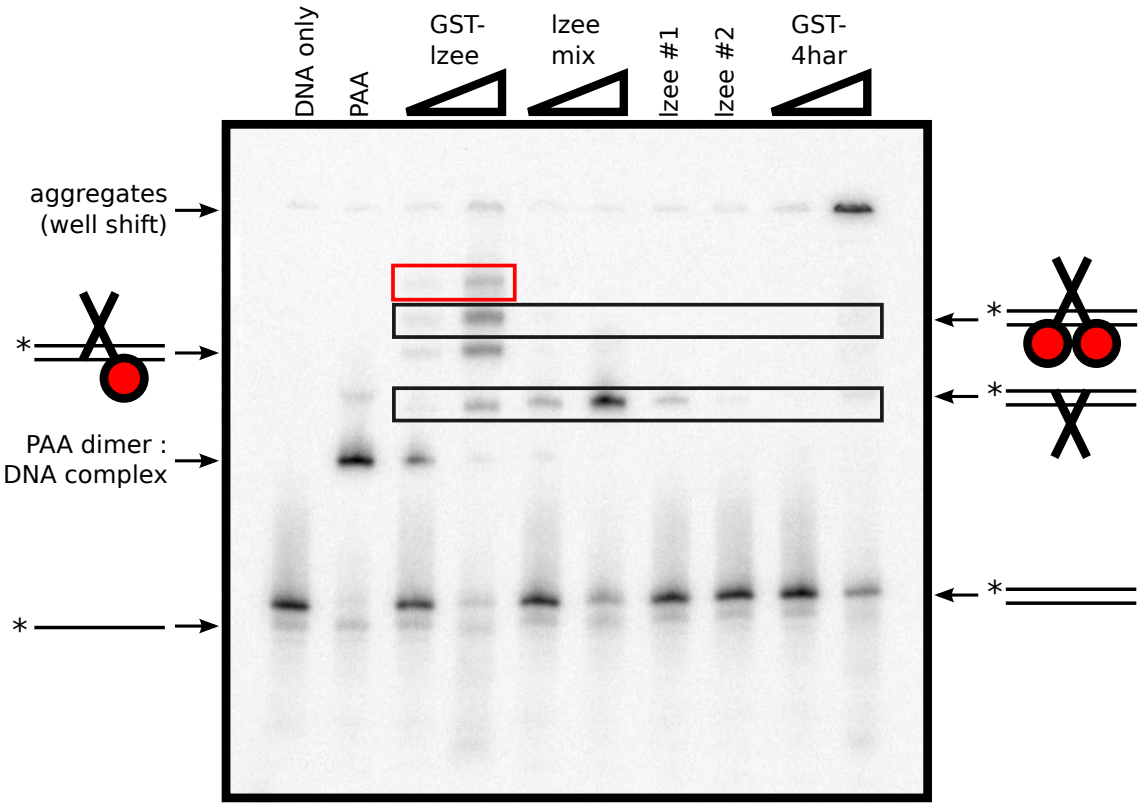


Figure 2.6: EMSA on GST-4har and cleaved off GST-lzee with 26-mer CREB DNA. The GST affinity tags are shown as red circles. ^{32}P-labeled DNA in each binding reaction is 25 nM. PAA is a GCN4 peptide derivative that is used as a positive control for protein dimer binding to CREB DNA. Different shifted bands in the (partially degraded) GST-lzee are due to the different combinations of protein dimers bound to the DNA. The affinity tag was cleaved off using PreScission protease (lzee mix), and the untagged protein was subsequently isolated by FPLC (lzee #1). In-column cleavage of the GST tag was also done, which led to a significant loss of the untagged protein yield (lzee #2). The similarity in band shifting behavior seen in PAA and untagged lzee indicate that lzee can not bind to two DNA fragments and form a sandwich complex. The additional band shift (in red box) upon GST-lzee binding is most likely a result of non-specific binding at higher protein concentrations.

2.5 Conclusions

Based on the results presented, we have achieved a few goals. First, we have successfully expressed and purified two of the LZT proteins, 4har and lzee, by fusing them to a GST affinity tag, and producing them in *E. coli* cells. Second, we have

shown that 4har forms aggregates with itself along with the DNA it is bound to. Lastly, we have illustrated that lzee can only bind to DNA as a dimer. It is important to note that the initial characterizations presented so far are done *in vitro*. The fact that initial cloning attempts of these proteins had deleterious effects in the host cells led us to the possibility that these proteins may be able to form tetramers, as originally designed, and stabilize *in vivo* DNA loops. Thus, we have adapted a few more experiments that would study the possible looping abilities of the LZT proteins *in vivo*.

Chapter 3: Probing the *in vivo* looping efficiencies of LZT and LZD proteins

3.1 Objectives

In vitro characterization of the putatively tetrameric protein lzee using EMSA showed no evidence of a protein-DNA sandwich complex, which is a necessary precursor for forming DNA loops. On the other hand, the LZD proteins had been previously shown to loop DNA *in vitro* (Gowetski et al., 2013). In this chapter, we adapted a quantitative assay to probe the looping ability of lzee, and the LZD proteins, *in vivo*. This assay incorporates the elements of the *lac* operon assembled together to generate a repression assay dependent on DNA looping. Furthermore, we tested various indicator media as a qualitative means to screen for transformants with downregulated reporter protein activity, to complement the quantitative repression assay. The main goals for this set of experiments are to (1) determine the smallest stable loops that can be mediated by our artificial DNA-binding proteins *in vivo*, (2) show how the length and orientation of the binding sites in the three LZD proteins affect the degree of reporter protein repression, which in turn characterize *in vivo* DNA flexibility, and (3) to combine the use of both the colorimetric screen

and the quantitative *in vivo* repression assay as a toolkit to design more efficient DNA looping proteins.

3.2 Rationale

The repression-based genetic assay relies on the concept of repression by "co-operativity at a distance" (Peters et al., 2011). In this assay, the F' episome in specialized *E. coli* cells was modified to place the *lacZYA* genes under the control of the *lacUV5* promoter (Whipple, 1998). As illustrated in Figure 3.1, a primary operator site, to which a DNA-binding protein can bind, is located downstream of the -10 region of the *lacUV5* promoter, in the vicinity of the +1 start site for the transcription of *lacZYA*. When the DNA binding protein is expressed, it binds to the primary operator in the F' episome, which represses the expression of the *lacZYA* genes. If the DNA binding protein contains two DNA-binding domains, a second operator located upstream of the -35 region of the *lacUV5* promoter will enhance reporter gene repression via protein delivery by looping of the intervening DNA.

To generate these specialized reporter strains, we modified several plasmid-based reporter constructs used in a previous study by Maher and colleagues (Becker et al., 2005), which were then introduced into the F' episome by homologous recombination. These constructs contain *cis* elements derived from the *lac* operon – the gene sequence that codes for the 250 amino acids at the C-terminal end of the lactose repressor (lacI), a distal operator O_d located at varying distances upstream of the *lacUV5* promoter, a proximal operator O_p responsible for repression, and

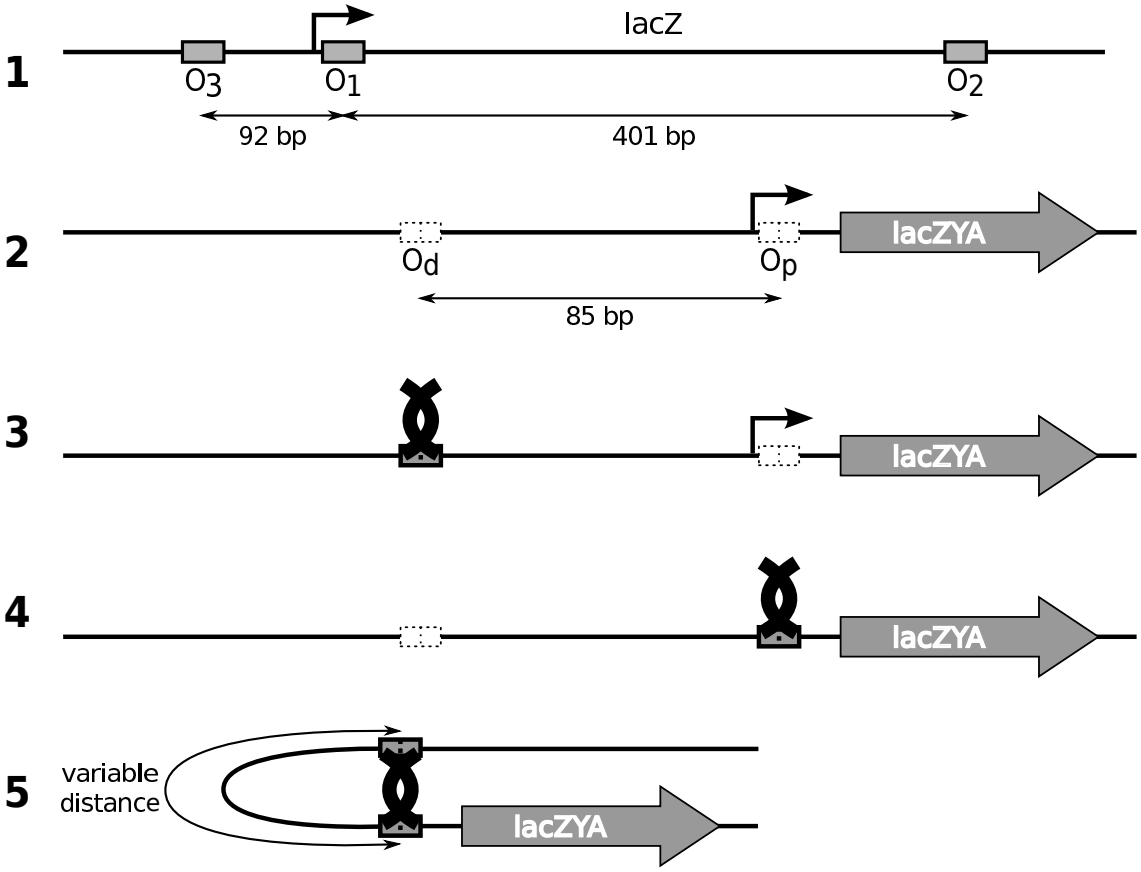


Figure 3.1: *In vivo* repression assay, adapted from Becker et al. (2005). Reporter strains were prepared, containing *cis* elements from the *lac* operon (1) in single copy via homologous recombination to the F' episome. In the recombinant F' episome, the O_2 operator was inactivated by mutagenesis, while the O_3 and O_1 sites were replaced with binding sites for the looping protein of interest at the distal (O_d) and proximal (O_p) positions, respectively. In order to detect looping-dependent repression of the the *lacZYA* genes, the following criteria must be met: Reporter gene expression should not be affected by looping protein binding to reporter strains that do not contain binding sites (2), or O_d only binding sites (3); There should be weak repression upon looping protein expression in reporter strains containing a low-affinity O_p only (4); In the presence of a high-affinity O_d (5), *lacZYA* repression would be more enhanced due to the tethering of the looping protein at the distal site, which increases the looping protein's local concentration to the primary proximal site.

codons 8-212 of the β-galactosidase (*lacZ*) gene. In addition, the kanamycin resistance gene is cloned in between *lacI* and the operator-promoter regions. Our looping constructs contained the high-affinity palindromic GCN4 operator ATF/CREB (5'-

ATGACGTCAT-3') (Hill et al., 1986) or the inverted INV2 site (5'-GTCATATGAC-3'), recognized by a C-terminal *reverse*GCN4 protein (Hollenbeck et al., 2001), at various distances in the distal site, with either the ATF/CREB or the wild-type GCN4 operator (5'-ATGACTCTT-3') placed in the proximal position (Hill et al., 1986). In addition, ATF/CREB and wild-type GCN4 sites are each cloned at the upstream and downstream positions separately, as non-looping constructs that can serve as negative and positive controls for *lacZYA* repression, respectively. The combination of operators has to be optimized such that the decrease in reporter protein levels would be more sensitive to the increased local concentration of the looping protein to the primary operator via DNA looping than to random binding of the looping protein to its primary site. Figure 3.2 lists the various promoter-operator regions used in the initial characterization for binding and looping of our putative looping proteins.

The *in vivo* assay also requires regulated amounts of the DNA-binding protein such that the amounts of the putative protein are high enough to affect reporter protein levels, but not detrimental to the host cell's growth. However, most expression plasmids are designed for large-scale expression and purification of the protein of interest, which is subsequently used for *in vitro* studies. Using these expression plasmids as-is would saturate the system with the protein being studied, which would mask the effect of DNA looping on reporter protein repression. Two strategies were used to adjust protein expression. First, the *tac* promoters controlling lzee and LZD expression in pGEX6P1 were changed into *lacUV5* or wild-type *lac* promoters, which adjust the amounts of the expressed proteins to levels more appropriate

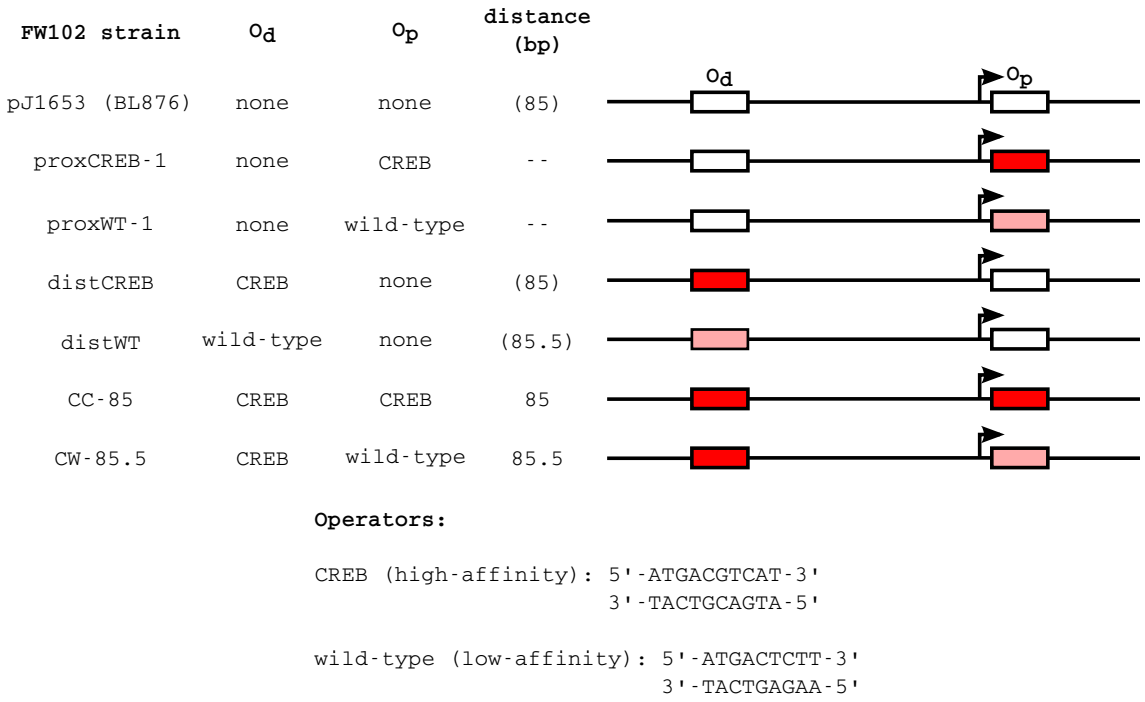


Figure 3.2: Promoter-operator regions of the reporter (FW102) strains used for the optimizing operators. GCN4 binding sites CREB (red boxes) or wild-type GCN4 (pink boxes) were cloned into the distal (O_d) and proximal (O_p) positions. *LacI* and kanamycin resistance genes are located upstream of O_d, while *lacZYA* genes are located downstream of O_p.

in the proposed reporter assay. The other strategy was to subclone the genes for the putative looping proteins into pBAD, which allows for tighter regulation of the recombinant protein expression by the arabinose promoter.

Reporter protein activity was monitored qualitatively using available indicator media, and quantitatively in liquid cultures using a colorimetric assay based on the cleavage reaction of o-nitrophenyl-β-galactoside (ONPG). We have tested three types of indicator media – LB plates with the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), MacConkey agar, and tetrazolium agar. X-gal is a lactose analog, that acts as a substrate but not as an inducer

for β-galactosidase. Enzyme-catalyzed hydrolysis of X-gal yields galactose and a substituted indole, which dimerizes and is oxidized to form the highly insoluble blue dye indigo. Cells that have a fully functional β-galactosidase (Lac^+ strains) produce blue colonies in media containing X-gal, while those that produce very low levels or mutated β-galactosidase yield white or light blue colonies (Lac^- strains). MacConkey and tetrazolium agar, combined with a sugar source such as lactose, are rich media that are used to detect the degree of metabolism of the sugar source. MacConkey lactose media was initially formulated to selectively grow Gram-negative bacteria and detect lactose fermenters (MacConkey, 1908). It is a rich medium that contains bile salts and crystal violet to prevent growth of Gram-positive bacteria, and neutral red, which acts as a pH indicator. Lac^+ strains metabolize lactose and release acid into the immediate environment, thus colonies rendered are stained red and usually surrounded with a hazy red precipitate of bile salts, whereas weakly Lac^+ strains produce white or pale pink colonies. Tetrazolium agar contains the redox indicator triphenyl tetrazolium chloride (TTC), which detects metabolically active cells. Cells that grow in this medium can reduce TTC into a red insoluble compound triphenylformazan (TPF). When lactose is added into the medium, Lac^+ strains can utilize the added sugar and generate acid into the immediate environment due to fermentation. Excess acid prevents the formation of the formazan dye, hence colonies of lactose fermenters remain white, while weakly Lac^+ or Lac^- colonies are pink or dark red respectively (Shuman & Silhavy, 2003).

ONPG is another analog of lactose which can act as a substrate for β-galactosidase, but not as an inducer for lacZYA expression. Hydrolysis of ONPG

produces galactose and o-nitrophenol, and the latter can be quantified by absorbance at 420 nm. The rate of product formation is proportional to the amount of the reporter protein expressed in the host cells, which is the definition of a Miller unit.

3.3 Materials and Methods

The plasmids pJ1645, pJ1654, and pJ1655 used to make the promoter-operator constructs were a kind gift from L. James Maher III. The pBAD/Myc-HisA plasmid (Invitrogen) was provided by the Cropp lab. *E. coli* XL1-Blue and XL10-Gold cells used to propagate the plasmids were purchased from Stratagene. MacConkey Lactose medium and Antibiotic Medium 2 were purchased from Difco Laboratories while 2,3,5-triphenyl-2H-tetrazolium chloride (TTC) and o-nitrophenyl-β-D-galactoside (ONPG) were purchased from USB Corporation. Instructions for preparing the LB liquid media and all selection plates are given in Appendix 1. Glucose (1%), lactose (1%), IPTG (0.1 mM), Bluo-gal (100 μg/mL) or X-gal (0.5 mM), and the antibiotics kanamycin (50 μg/mL), chloramphenicol (50 μg/mL), ampicillin (100 μg/mL), and streptomycin (100 μg/mL) were added as indicated. Absorbance readings were done on a Cary 300 UV-vis Spectrophotometer.

3.3.1 Promoter-operator constructs

The first set of plasmid-based constructs containing one or two operators were generated using the non-looping (pJ1645 and pJ1654) and looping (pJ1655) plasmids, respectively as templates for PCR. The template plasmids contained the κB

operators. The distance between the GCN4 binding sites in the resulting looping constructs is 85 bp for two CREB sites, or 85.5 bp for one CREB site and one wild-type GCN4 site.

3.3.1.1 Single operator controls

The plasmids proxWT-1 and proxCREB-1were constructed by mutating the proximal operator, O_p, of pJ1645 into either the GCN4 wild-type (5'-ATGACTCTT-3'), or ATF/CREB (5'-ATGACGTCAT-3') operator using the QuikChange II XL Mutagenesis Kit (Stratagene), as shown in Figure 3.3. The resulting linear amplification products were digested with DpnI to remove parental DNA, then transformed into XL10-Gold or XL1-Blue cells, and grown on LB agar plates containing kanamycin and chloramphenicol. Transformants containing the mutated O_p sites gave a 151-bp PCR product using primers (a) Od-fp2.1 (5'-GGCGTATCACGAG GCCCTTTCGTCTTCAAGAATTC-3') and Op-rp2a (5'-GTGCGGCCGCGGTT GTGGGAATGACGT-3') to verify the proximal CREB site, or (b) Od-fp2.1 and Op-rp2b (5'-GTGCGGCCGCGGTTGTGGGAAAGAGT-3') to verify the proximal wild-type GCN4 site. The sequences of the correct transformants were then determined by DNA sequencing using seqFP-OdOp (5'-GGCCCGGAGGGTGGCGGG CAGG-3') as the sequencing primer.

Figure 3.4 shows the cloning strategy for plasmid distWTby introducing the wild-type GCN4 operator into O_d of pJ1654 using PCR with phosphorylated primers. The 5' ends of the resulting amplification product were ligated together to form the

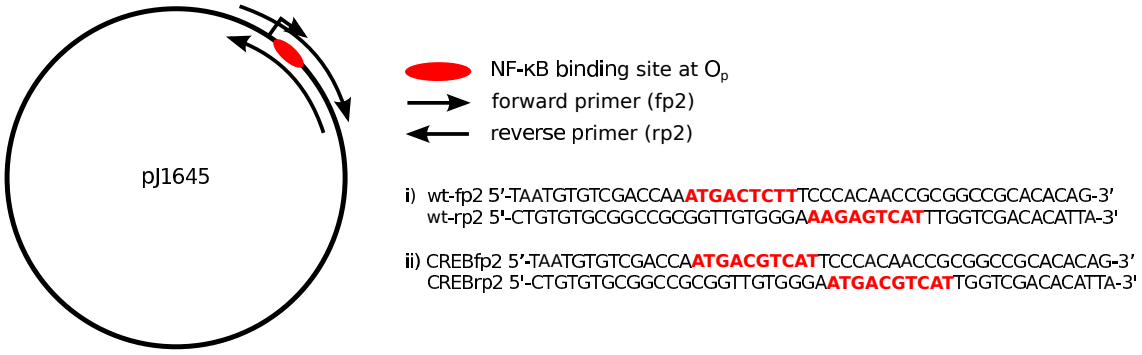


Figure 3.3: Cloning strategy for generating proxWT-1 and proxCREB-1, using site-directed mutagenesis. Complementary mutagenic primers containing the desired mutations are used for linear amplification of the plasmid template. The amplification product is a circular product with two single nicks on opposite strands, which is then transformed into the host cell for plasmid propagation.

circular plasmid. This ligation product was subsequently transformed into XL10-Gold or XL1-Blue cells and grown in LB agar plates containing kanamycin and chloramphenicol. Transformants were screened for the presence of the correct distal operator by *Bsr*GI digestion and verified by DNA sequencing with Op-rp1.1 (5'-T CATAGCTGTTTCCTGTGTGCGGCCGCG-3'). Cloning of the new operators at O_d introduced a single *Bsr*GI site between the new operator and the *Sac*I recognition site. This cloning maneuver enables easy switching of O_d while keeping the linker sequence between operators intact, or extending the linker sequence without affecting the O_d sequence.

The preparation of plasmid distCREB is illustrated in Figure 3.5. In brief, the ATF/CREB operator was cloned into the distal site by ligating the *Eco*RI-*Sac*I double digests of pJ1654 to the EcoRI-SacI double digest of the complementary mutagenic oligonucleotides that were annealed in the reaction buffer prior to enzymatic digestion. The crude ligation mixture was transformed into XL1-Blue cells

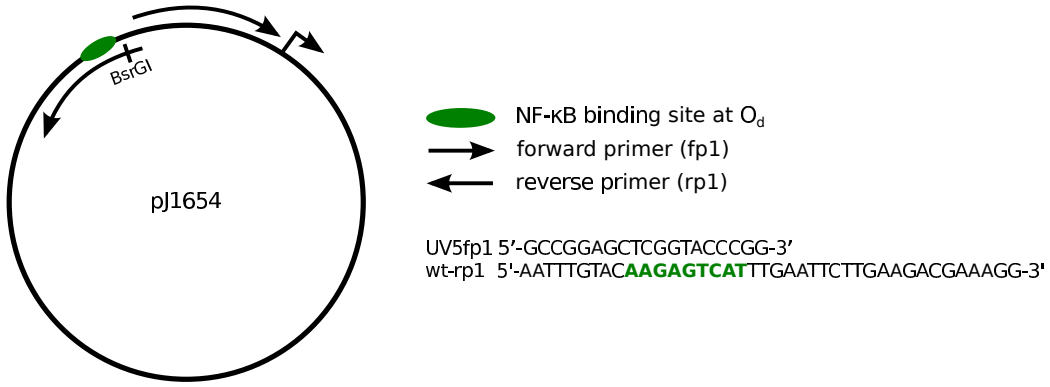


Figure 3.4: Cloning strategy for generating distWT, using phosphorylated primers. Mutations are introduced on one of the primers, which are then used for amplifying the plasmid template. The resulting product is circularized and transformed into the host cell for plasmid propagation

and grown in LB agar plates with kanamycin and chloramphenicol. Transformants were screened by *Bsr*GI digestion and verified by DNA sequencing using seqRP-OdOp (5'-GGATTGACCGTAATGGGATAGG-3') as the sequencing primer.

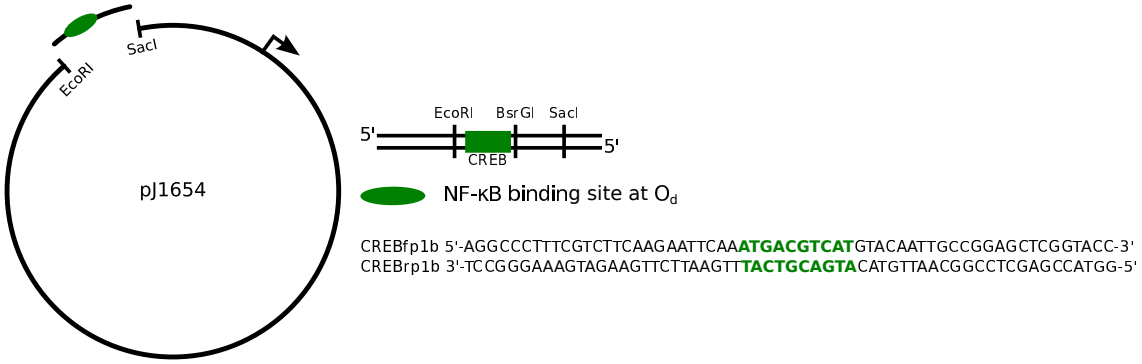


Figure 3.5: Cloning strategy for generating distCREB, using a simple cut-and-paste method. Complementary oligonucleotides containing the CREB site are annealed together, and digested with *Eco*RI and *Sac*I. The oligonucleotide insert is then cloned into the *Eco*RI – *Sac*I digestion product of the parent plasmid, and transformed into the host cell for plasmid propagation.

3.3.1.2 85- and 85.5-bp looping constructs

Looping constructs CC-85 and CW-85.5 were prepared in two steps, as shown in Figure 3.6. First, the ATF/CREB site was cloned into the distal site of pJ1655 using the same method in the previous paragraph. The correct transformant, CREB/pJ1655, was then used as the starting template in the next step. The GCN4 site was introduced in the distal site of promoter/operator region of the CREB/pJ1655 plasmid by amplifying with mutagenic primers (i) CREBfp1a and CREBrp2, or (ii) CREBfp1a and wt-rp2, generating a 141-bp PCR product containing (i) two ATF/CREB sites, or (ii) a distal ATF/CREB site and a proximal wild-type GCN4 site. The PCR products were then digested with *Eco*RI and *Not*I and cloned into the *Eco*RI-*Not*I double digest of pJ1655. The crude ligation mixtures were subsequently transformed into XL1-Blue cells and grown in LB agar plates containing kanamycin and chloramphenicol. Transformants containing the desired operators were then sequenced using Od-fp2.1 and Op-rp1.1 as the sequencing primers.

3.3.1.3 138.5- to 237.5-bp looping constructs with proximal wild-type GCN4 site

The second set of promoter-operator constructs were prepared from CW-85.5. The separation between the GCN4 sites was increased to 138.5, 187.5 or 237.5 bp. Figure 3.7 illustrates how the linker sequences were generated. PCR amplification

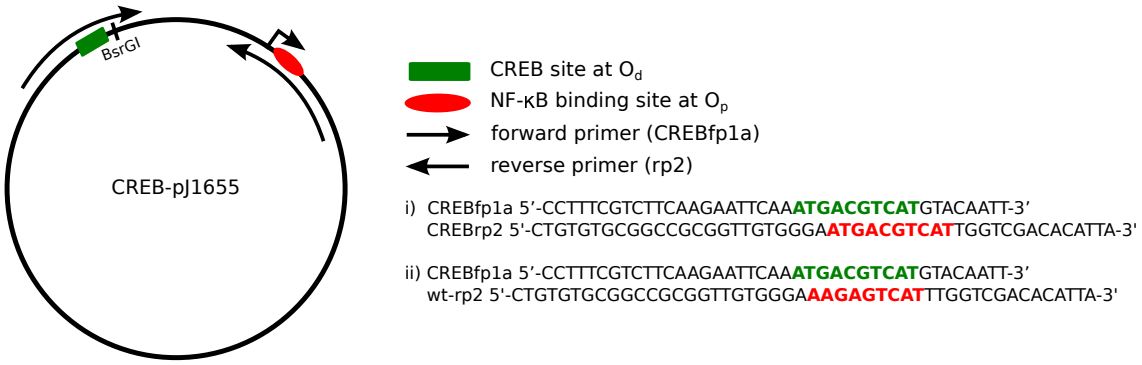


Figure 3.6: Cloning strategy for generating looping constructs, 85- or 85.5-bp distance between operators. The CREB site is cloned into the plasmid as described in Figure 3.5. Then, the second operator is introduced into the downstream site by PCR. The resulting PCR product is digested with *Eco*RI and *Not*I, and ligated into the *Eco*RI – *Not*I digest of the parent plasmid. The ligation product is then transformed into the host cell for plasmid propagation.

from the pRSETA plasmid using the primer pairs (a) SMSlinkFP + SMSlinkRP1, (b) SMSlinkFP + SMSlinkRP2, and (c) SMSlinkFP + SMSlinkRP3, generated a 85-, 139-, and 187-bp PCR product, respectively. Both the parent plasmid (CW-85.5) and aforementioned PCR products were digested with *Sac*I and *Bam*HI, with the parent plasmid dephosphorylated. The resulting digestion products are then ligated together, transformed into XL1-Blue cells, and grown on LB plates containing kanamycin and chloramphenicol. Correct transformants gave a 203-, 252-, and 302-bp PCR product using primers Od-fp1a.1 (5'-CGTATCACGAGGCCCTTTC GTCTTCAAGAATTCAAATGAC-3') and Op-rp2b. The plasmid minipreps were then sequenced using seqFP-OdOp as the sequencing primer. The sequence-verified plasmids were identified as MP56-14a, MP56-15c and MP56-16a, which we later renamed to CW-138.5, CW-187.5, and CW-237.5, respectively.

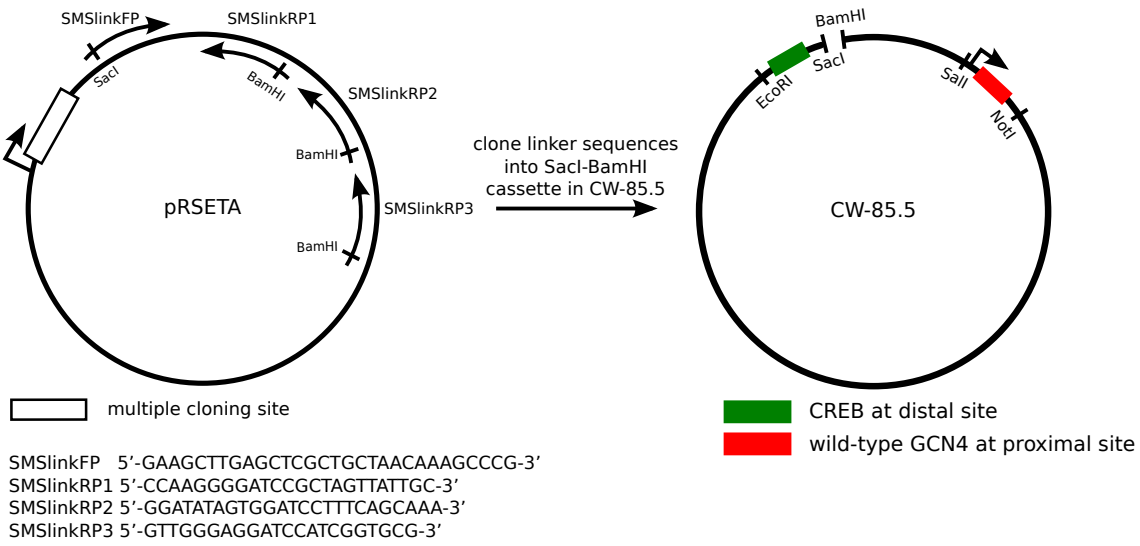


Figure 3.7: Cloning strategy for generating looping constructs, 138.5- to 237.5-bp distance between operators. The locations of the reverse primers are not drawn to scale.

3.3.1.4 300- to 1150-bp looping constructs with distal INV2 site

The last set of promoter-operator constructs were prepared using proxCREB-1 as parent plasmid. A 1120-bp linker sequence, which was PCR-amplified from lambda phage using the primers lambdaFP-EcoRI and lambdaRP-SacI and digested with *Eco*RI and *Sac*I, was cloned into the *Eco*RI-*Sac*I region in proxCREB-1. Correct clones were verified by colony PCR using seqFP-OdOp and Op-rp2a primers, which gave a 1382-bp product. We have named this plasmid proxCREB-2. Then, a second symmetrical GCN4 binding site, INV2 (5'-GTCATATGAC-3') was cloned at various upstream positions (300, 450, 650, 900, and 1150 bp away from CREB) by mutagenesis PCR similar to the preparation of distWT plasmid (see Figure 3.4). The correct clones were verified by *Bsr*GI or *Spe*I digestion, and sequenced using seqFP-OdOp as sequencing primer. We have named the correct transformants

IC-300, IC-450, IC-650, IC-900 and IC-1150. Finally, the IW series of plasmids (IW-300.5, IW-450.5, IW-650.5, IW-900.5 and IW-1150.5) were generated by switching out the CREB operators in the IC plasmid series with the wild-type GCN4 binding sites. Primer sequences for the linker sequences and the mutagenesis primers for cloning in INV2 sites are listed in Appendix 2.2.

All promoter-operator constructs were then sent to the Maher lab for recombination experiments, generating our specialized *E. coli* reporter strains. The full sequences of the promoter-operator constructs are included in Appendix 2.3.

3.3.2 Expression plasmids

The following phosphorylated primers were used to mutate the *tac* promoter in the pGEX6P1 plasmids to (a) *lacUV5* and (b) *lacWT*: (a) mutFP-lacUV5 (5'-TTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGG-3') and mutRP-pGEX (5'-CAGCTCATTTCAGAATATTTGCCAGAACCGTTATGATGTCGGCGC-3'), (b) mutFP2-pGEX (5'-GTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGTATTCATGTCCCCTATAC-3') and mutRP2-lacWT (3'-GGCTGTAGTATTGCCAAGACCGTTTATAAGACTTTACTCGACAAATGTGAAATACGAAGGCCGAGCATACAA-5'). The PCR products were ligated, transformed into XL-1 Blue cells and grown in LB plates containing ampicillin. Transformants were screened for the presence of the mutated promoter by *Pvu*II digestion, and verified by DNA sequencing with seqFP2-pGEX (5'-CGTTCTGGATAATGTTTTTTGCGCC-3').

To subclone the DNA-binding proteins in pBAD/Myc-HisA, the gene inserts in the recombinant pGEX6P1 plasmids were amplified by PCR using the following primer pairs: (a) pGEXfp2 (5'-CCAAAATCGGGTACCGAAGTTCTGTTCC-3') and pGEXrp (5'-CGATGCGGCCGCTCGAGTC-3'), and (b) pGEXfp3 (5'-ACA GGAAACAGGTACCATGTCCCCTATACTAGGTTATTGG-3') and pGEXrp. Primer pair (a) amplified the gene inserts without the affinity tag, while the pair (b) amplified the genes with the N-terminal GST tag. The resulting PCR products were then digested with *Kpn*I and *Eco*RI, purified by native PAGE, and subsequently subcloned into the *Kpn*I-*Eco*RI digestion product of pBAD/Myc-HisA plasmid. Recombinant plasmids were then identified using PCR and verified by DNA sequencing.

The sequences of the expression plasmids used in this study are included in Appendix 2.1.

3.3.3 Reporter protein assays

The expression plasmids (recombinant pGEX6P1 or pBAD/Myc-HisA) were transformed into electrocompetent reporter strains using standard electroporation protocols, and grown on selection plates containing ampicillin, kanamycin, and streptomycin. The following selection plates were used in the course of characterizing the putative DNA binding proteins: (1) LB, (2) LB X-gal or Bluo-gal ± IPTG, (3) MacConkey lactose ± IPTG, (4) tetrazolium glucose, and (5) tetrazolium lactose. Individual colonies were selected for starter culture growth overnight (∼16 hours)

at 37 °C in LB supplemented with amp, kan and strep. The starter cultures were then diluted 50-fold with fresh growth media varying concentrations of IPTG (0 – 1 mM) or arabinose (0 – 0.2%) and incubated at 37 °C with agitation for two hours or until A_{600} ∼0.3 – 0.7. For samples with high β-galactosidase levels, 100-μL aliquots of the resulting cultures were mixed with 900 μL Z-buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol). Cell cultures with low β-galactosidase levels were pelleted out from 1-mL aliquots, and resuspended in 1 mL Z-buffer. The cells were then lysed by pulse-vortexing for 10 seconds with 50 μL chloroform and 25 μL 0.1% SDS. The lysed cells were equilibrated at 30 °C for 5 minutes, then the substrate for the assay (200 μL of 4 mg/mL o-nitrophenyl-β-galactoside in Z-buffer) was added and the mixtures were incubated for 10 – 15 minutes at 30 °C. The reactions were subsequently quenched by adding 500 μL of 1 M Na_2CO_3 and reaction times were recorded. Absorbance readings at 420 nm and 550 nm were then measured from 1-mL aliquots of the reaction mixtures, and at 600 nm from 1-mL aliquots of the working cultures. Raw enzyme units (E) were then calculated using the formula: 1000 x [A_{420} - 1.75 x A_{550}]/[reaction time (in minutes) x culture volume (in mL) x A_{600}] (Miller, 1972). The repression ratios for all looping and non-looping constructs were calculated as RR = E_{-IPTG}/E_{+IPTG}, where E_{+IPTG} is the raw enzyme units for transformants induced at a specific IPTG concentration and E_{-IPTG} is the corresponding raw enzyme units for uninduced transformants. Two replicates per transformant in a selection plate were analyzed.

3.4 Results and Discussion

3.4.1 Baseline expression in reporter strains

Baseline levels of the β-galactosidase in all of the reporter strains studied were first determined qualitatively and quantitatively. The reporter strains were plated onto LB plates containing antibiotics and Bluo-gal. All reporter strains that have intact *lacUV5* promoter in the F' episome would fully express the reporter protein, and hydrolyze Bluo-gal to stain colonies blue. All of the reporter strains in this study yielded blue colonies, indicating that the recombination procedures were successful. Plating the same strains onto LB Bluo-gal plates containing IPTG also produced blue colonies, which would indicate that baseline amounts of β-galactosidase in untransformed reporter strains are not affected by IPTG. Reporter protein activities on liquid cultures of transformants from both selection plates have verified this observation (see Figure 3.8).

3.4.2 Testing operator strength with 4har and lzee proteins

Next, the reporter strains FW102/CC-85 and FW102/CW-85.5 were transformed with recombinant pGEX6P1 (4har and lzee), in which expression of the gene inserts is controlled by the *tac* promoter. We want to test whether over-expression of the tetrameric DNA-binding proteins would elicit a change of phenotype in the reporter strains. The resulting transformants produced blue colonies in both Bluo-gal and Bluo-gal + IPTG plates. Blue colonies were then selected from Bluo-gal

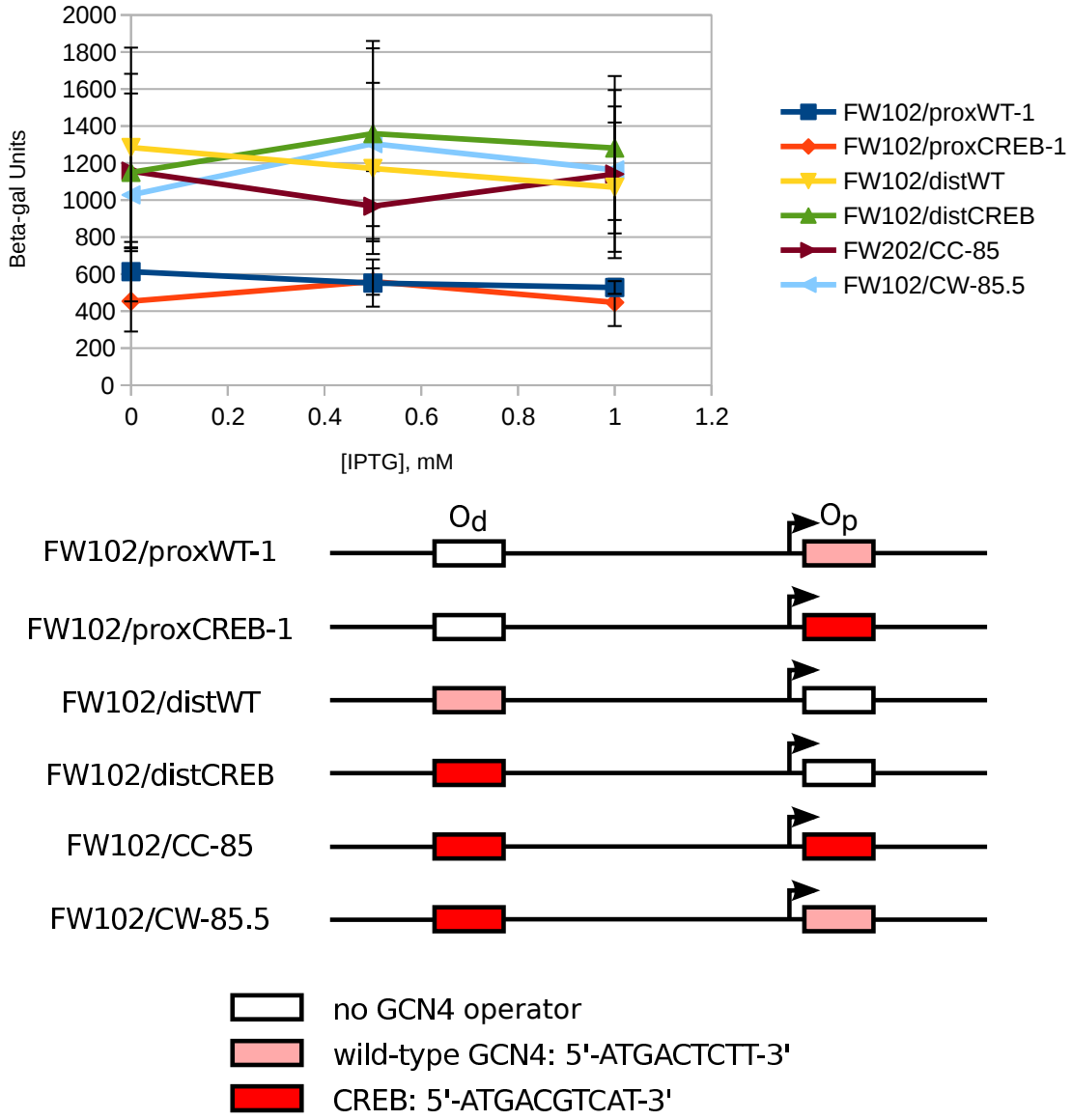


Figure 3.8: β-galactosidase activity in untransformed reporter strains. The values reported are average numbers from four replicates, in which two were selected from LB Bluo-gal plates and the other two from LB Bluo-gal IPTG plates. Results show that the reporter protein activities are not affected by IPTG

plates and analyzed for reporter protein activity (see Figure 3.9). Quantitative reporter protein assay results showed no significant repression in the FW102/CW-

85.5 reporter strain upon dose-dependent expression of either 4har or lzee protein. This correlates well with the phenotype generated from the Bluo-gal IPTG plate. On the other hand, FW102/CC-85 showed at least 50% repression in the presence of 4har, and 83% repression from lzee expression, although the measured values were no less than 242 (for 4har) or 44 (for lzee) β-gal units. However, a change in color from blue to white only occurs from loss of reporter protein activity, corresponding to $<$20 β-gal units (Whipple, 1998). Thus, the observed phenotypes for FW102/CC-85 transformants in the X-gal + IPTG plates still match with the measured β-galactosidase activities.

Blue colonies were also observed in both Bluo-gal and Bluo-gal IPTG plates after transforming the reporter strains containing proximal operators only (FW102/proxCREB-1 and FW102/proxWT-1) with 4har and lzee. Quantitative reporter activities also show similar trends as their corresponding looping strains (see Figure 3.10). However, in both cases, there is no significant difference in the repression levels due to the presence of the upstream GCN4 site. This suggests that the repression level of β-gal exhibited by looping strains is not due to DNA looping.

On the other hand, transformation of reporter strains containing GCN4 sites in the distal position (FW102/distCREB and FW102/distWT) with 4har or lzee expression plasmids generated a lawn of white colonies in selection plates containing Bluo-gal $\pm$ IPTG. It turned out that saturation of the selection plates with distal non-looping transformants could give false negative results , as restreaking white colonies onto fresh selection plates yielded individual blue colonies on both types of selection media. Moreover, quantitative reporter activities showed no dose-

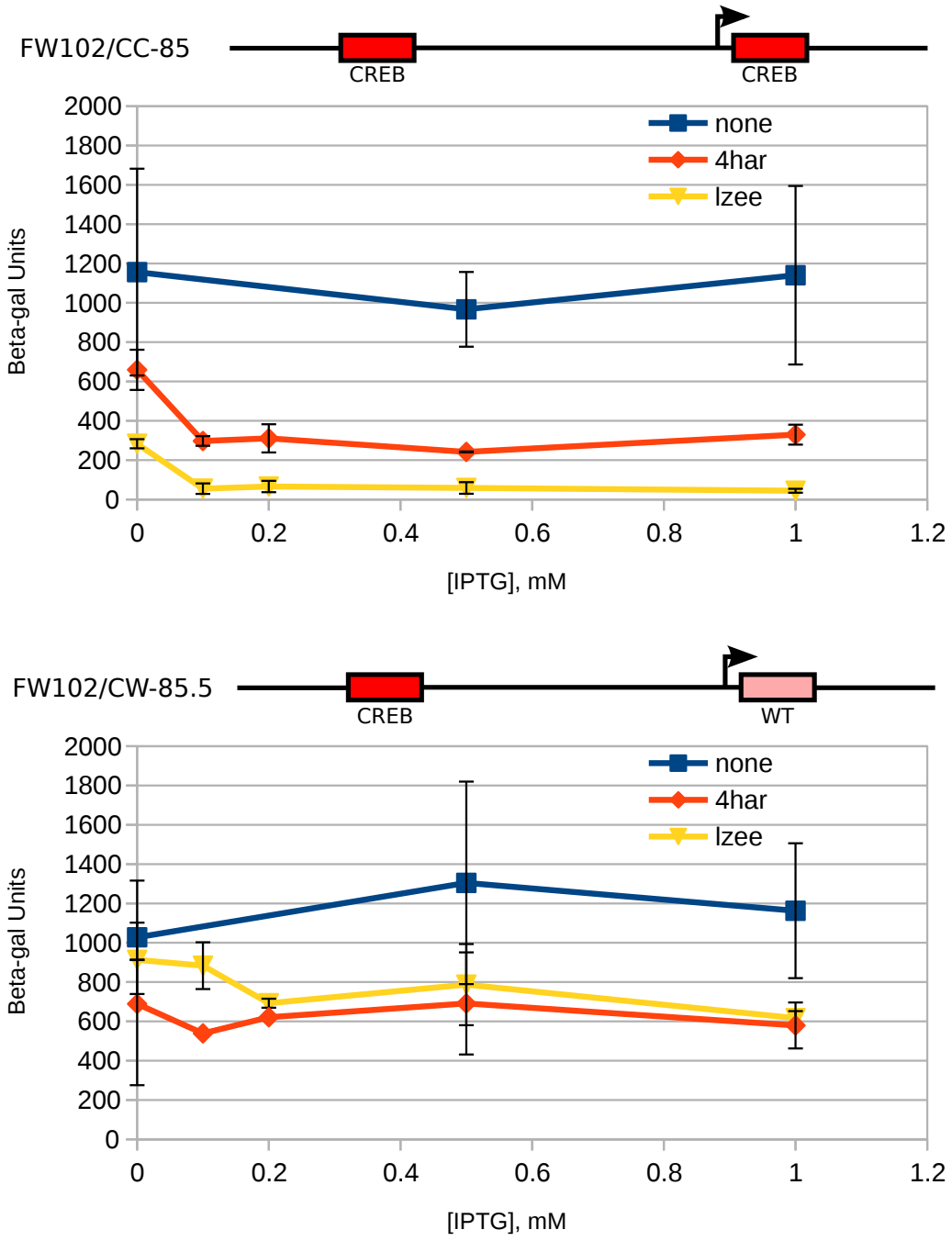


Figure 3.9: Comparing β-galactosidase activities in looping strains transformed with 4har and lzee, highlighting the difference in the operator strengths at the proximal site. Both proteins bind more strongly to the palindromic CREB site (5'-ATGACG TCAT-3') in FW102/CC-85 than the pseudopalindromic wild-type GCN4 site (5'-ATGACTCTT-3') in FW102/CW-85.5. The numbers reported are average values from two replicates selected from LB Bluo-gal plates.

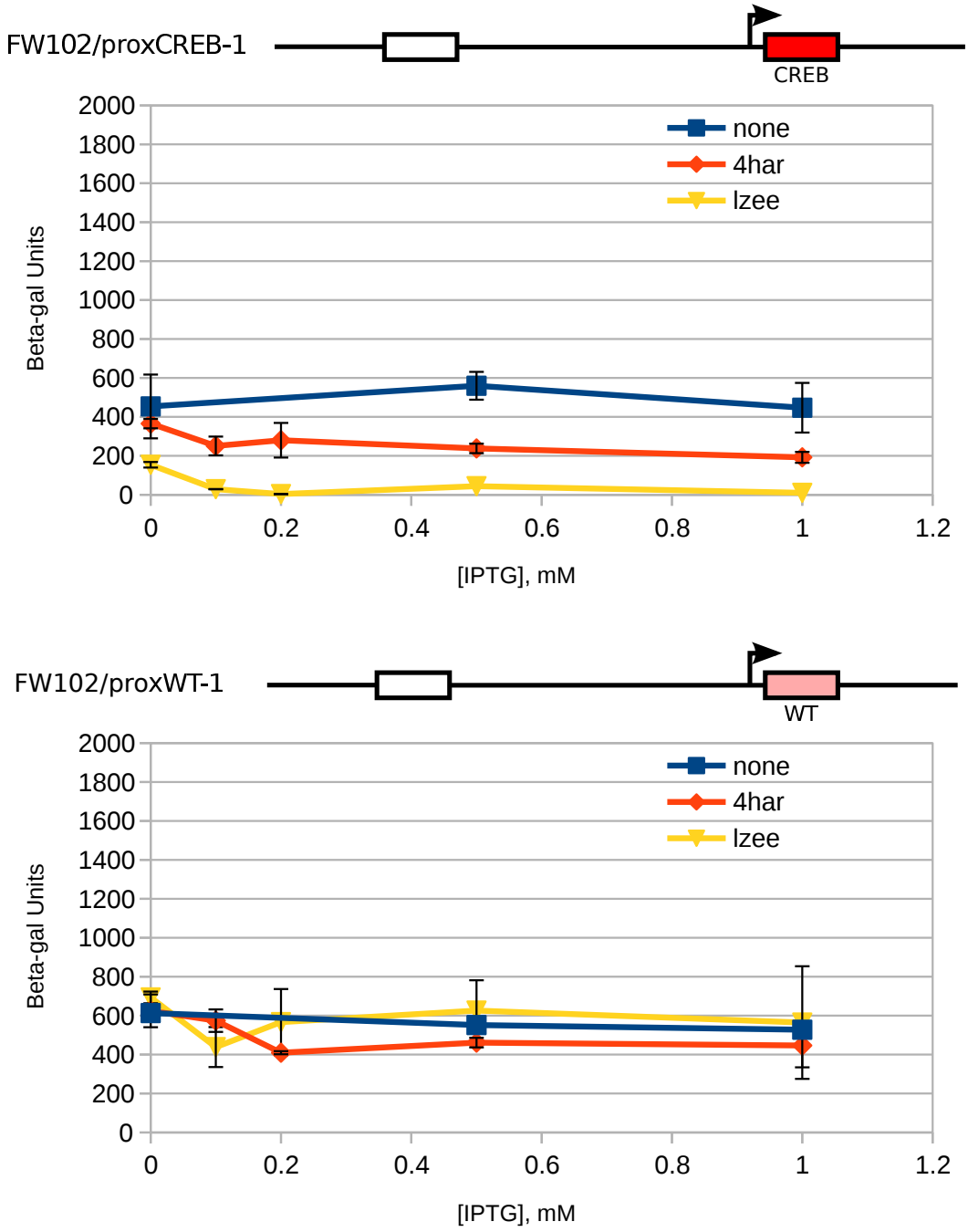


Figure 3.10: Comparing β-galactosidase activities in "O_p only" non-looping strains transformed with 4har and lzee, highlighting the difference in the operator strengths at the proximal site. Both proteins bind more strongly to the CREB site than the wild-type GCN4 site. The values reported are average numbers from two replicates selected from LB Bluo-gal plates.

dependent repression in any of these transformants upon IPTG induction, as illustrated in Figure 3.11.

The first set of results of the β-galactosidase assay, illustrated in Figures 3.9 – 3.11 are summarized in Tables 3.1 – 3.2. Several points can be derived from the data. First, in the absence of the protein expression plasmid, cell strains did not show significant dose-dependent repression of *lacZYA* upon induction with IPTG (RR values do not deviate much from 1.0). This is to be expected, as the promoter-operator constructs were designed such that the primary and auxiliary operators in the wild-type lac operon have been removed, and replaced with the GCN4 binding sites. Thus, the reporter protein in the F' episome can only be inhibited by our putative DNA-binding proteins.

Table 3.1: Reporter protein behavior in looping strains, with GCN4 sites 85 or 85.5 bp apart

Looping Strain	Expression Plasmid	β-gal Units (0 mM IPTG)	β-gal Units (1 mM IPTG)	Repression Ratio, RR
FW102/CC-85	—	1156 ± 526	1140 ± 454	1.0 ± 0.6
FW102/CC-85	tac-4har	659 ± 102	330 ± 51	2.0 ± 0.4
FW102/CC-85	tac-lzee	283 ± 23	44 ± 10	6 ± 2
FW102/CW-85.5	—	1028 ± 289	1163 ± 343	0.9 ± 0.4
FW102/CW-85.5	tac-4har	689 ± 413	579 ± 117	1.2 ± 0.8
FW102/CW-85.5	tac-lzee	913 ± 1	617 ± 36	1.48 ± 0.09

Second, there was no significant change in reporter gene levels for strains containing strong or weak operators in the distal (upstream) position, upon 4har or lzee expression. This suggests that a protein bound to the distal site has no effect on promoter strength, and any non-specific loops formed from the tethering of the

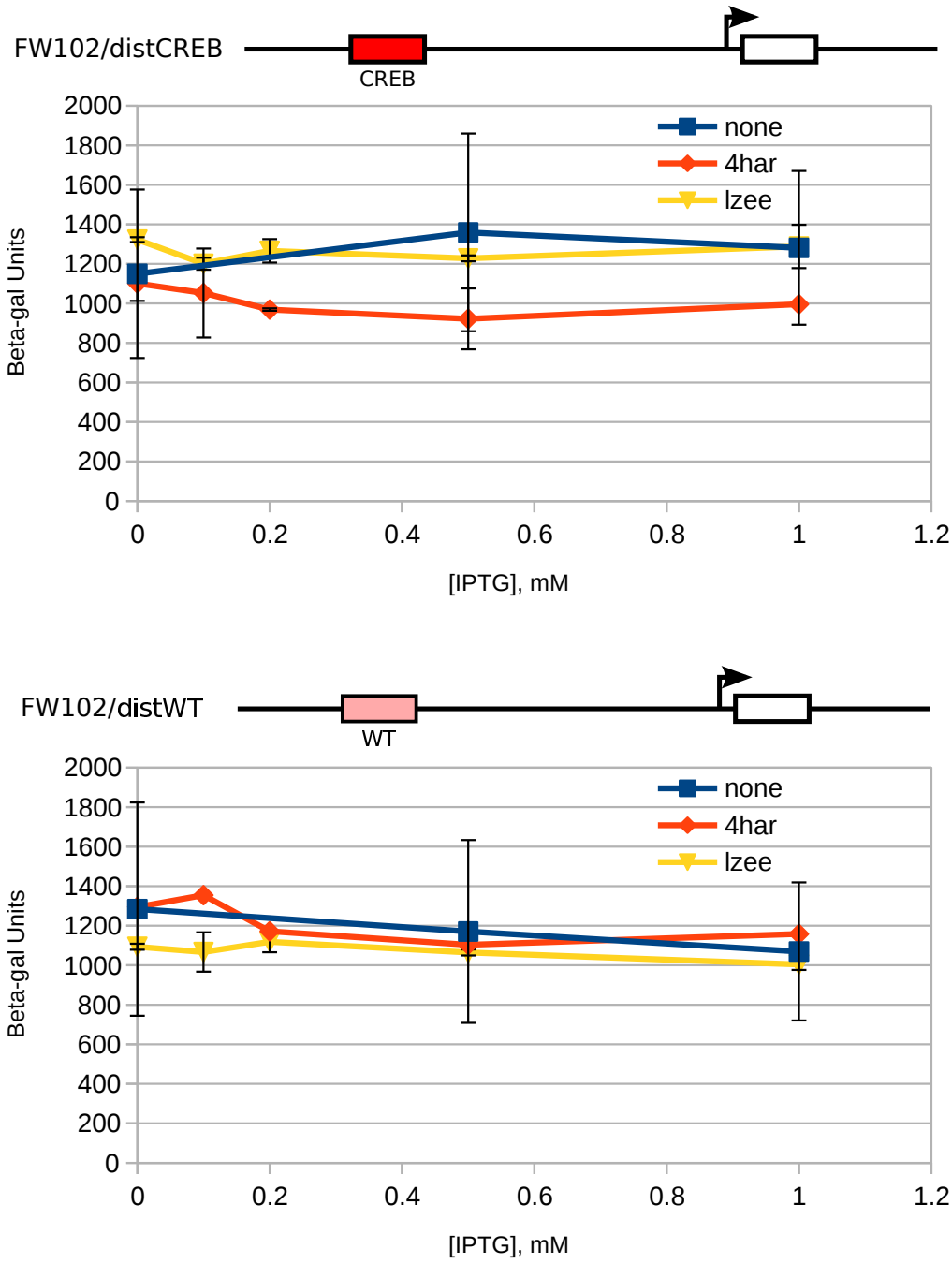


Figure 3.11: Comparing β-galactosidase activity in "O_d only" non-looping strains transformed with 4har and lzee, highlighting the effect of the distal site on repression. Repression levels are not affected by the presence of an upstream GCN4 site, signifying that any nonspecific loops formed upstream of the promoter do not repress the reporter protein. The values reported are average numbers from two replicates selected from LB Bluo-gal plates.

Table 3.2: Reporter protein behavior in "O_p only" non-looping strains. Only one tac-lzee:FW102/proxCREB-1 transformant gave positive reporter activity at 1 mM IPTG.

Proximal Operator	Expression Plasmid	β-gal Units (0 mM IPTG)	β-gal Units (1 mM IPTG)	Repression Ratio, RR
CREB	—	454 ± 164	447 ± 127	1.0 ± 0.5
CREB	tac-4har	365 ± 24	192 ± 28	1.9 ± 0.3
CREB	tac-lzee	154 ± 14	10	15
Wild-type	—	613 ± 161	527 ± 35	1.2 ± 0.3
Wild-type	tac-4har	625 ± 84	447 ± 112	1.4 ± 0.4
Wild-type	tac-lzee	693 ± 30	564 ± 289	1.2 ± 0.6

Table 3.3: Reporter protein behavior in "O_d only" non-looping strains. Only one tac-4har/Wild-type (GCN4) transformant was successfully analyzed for this experiment

Distal Operator	Expression Plasmid	β-gal Units (0 mM IPTG)	β-gal Units (1 mM IPTG)	Repression Ratio, RR
CREB	—	1150 ± 426	1281 ± 389	0.9 ± 0.4
CREB	tac-4har	1108 ± 88	996 ± 88	1.1 ± 0.1
CREB	tac-lzee	1323 ± 12	1288 ± 110	1.03 ± 0.09
Wild-type	—	1284 ± 540	1070 ± 349	1.2 ± 0.6
Wild-type	tac-4har	1295	1158	1.1
Wild-type	tac-lzee	1094 ± 15	1004 ± 28	1.09 ± 0.03

protein upstream of the promoter would not repress the reporter gene (Becker et al., 2005).

Third, there is significant repression at non-inducing conditions from lzee transformed into reporter strains containing CREB at the proximal site. On the other hand, the same expression plasmid does not have a similar effect on reporter strains containing the downstream wild-type GCN4 site. The baseline repression is most likely due to leaky expression of lzee, which has stronger affinity to the CREB site.

To summarize our findings so far, our initial reporter gene assay results indicate that lzee is a fairly good repressor but not an efficient looping protein on operators spaced 85.5 base pairs apart. The distance between the operators in the looping strains is ∼28.9 – 29.1 nm. Given that 4har and lzee are both ∼18.75 nm, as measured using PyMOL on the composite protein image generated from PDB files 2DGC (Keller, König, & Richmond, 1995) and 1TLF (Friedman, Fischmann, & Steitz, 1995), the intervening distance between the operators is too short to produce a stable DNA loop with the LZT proteins. Therefore, longer looping strains were made to address this issue. Moreover, having longer looping strains would enable us to further optimize the proximal operator such that it might significantly enhance repression upon looping protein expression if the strong distal operator is far enough upstream to form a stable loop.

3.4.3 Testing operator strength with LZD proteins

The three LZD variants (LZD73, LZD80, and LZD87), along with lzee and the empty vector as positive and negative controls for repression, respectively, were then transformed into FW102/CC-85 and FW102/CW-85.5 strains, and plated onto LB X-gal plates. All transformants yielded blue colonies. Streaking the blue colonies onto X-gal plates containing IPTG resulted in blue colonies with occasional white ones (∼1%). Initial protein expression and purification experiments done on the LZD proteins showed that leaky transcription of these proteins proved to be toxic to the host cells. One way to circumvent this was to narrow down the range of IPTG

concentrations for the *in vivo* repression assay from 0 – 0.01 mM in an attempt to minimize, if not prevent, cell toxicity of the LZD proteins. Furthermore, blue colonies growing on Xgal plates were selected for the reporter protein assays, as the (rare) white ones are most likely aberrant transformants.

Miller assay results on lzee and the LZD proteins transformed into three reporter strains are shown in Figure 3.12 and Table 3.4. Similar trends to those from 4har and lzee can also be seen from the data. First, there is no significant repression of β-galactosidase on any of the transformants that contain the wild-type GCN4 site at the proximal position (FW102/CW-85.5). On the other hand, transformants containing the proximal CREB site showed some repression upon inducing protein expression from the recombinant plasmids, but not from the empty vector (tac-pGEX6P1).

Among the expression plasmids analyzed, lzee and LZD87 transformed into reporter strains containing a proximal CREB site have relatively low levels of β-galactosidase at non-inducing conditions which decreased slightly in a dose-dependent manner upon protein expression. The rest of the FW102/CC-85 transformants have higher reporter protein activities, and exhibit dose-dependent β-gal repression in the range of IPTG used. Similar observations hold for all of the FW102/CW-85.5 transformants.

Finally, the recombinant pGEX plasmids were transformed into the BL876 reporter strain. The F' episome of BL876 does not contain any GCN4 sites in the promoter-operator region, thus reporter protein activity from this strain should not be affected by protein expression from the recombinant pGEX plasmids. However,

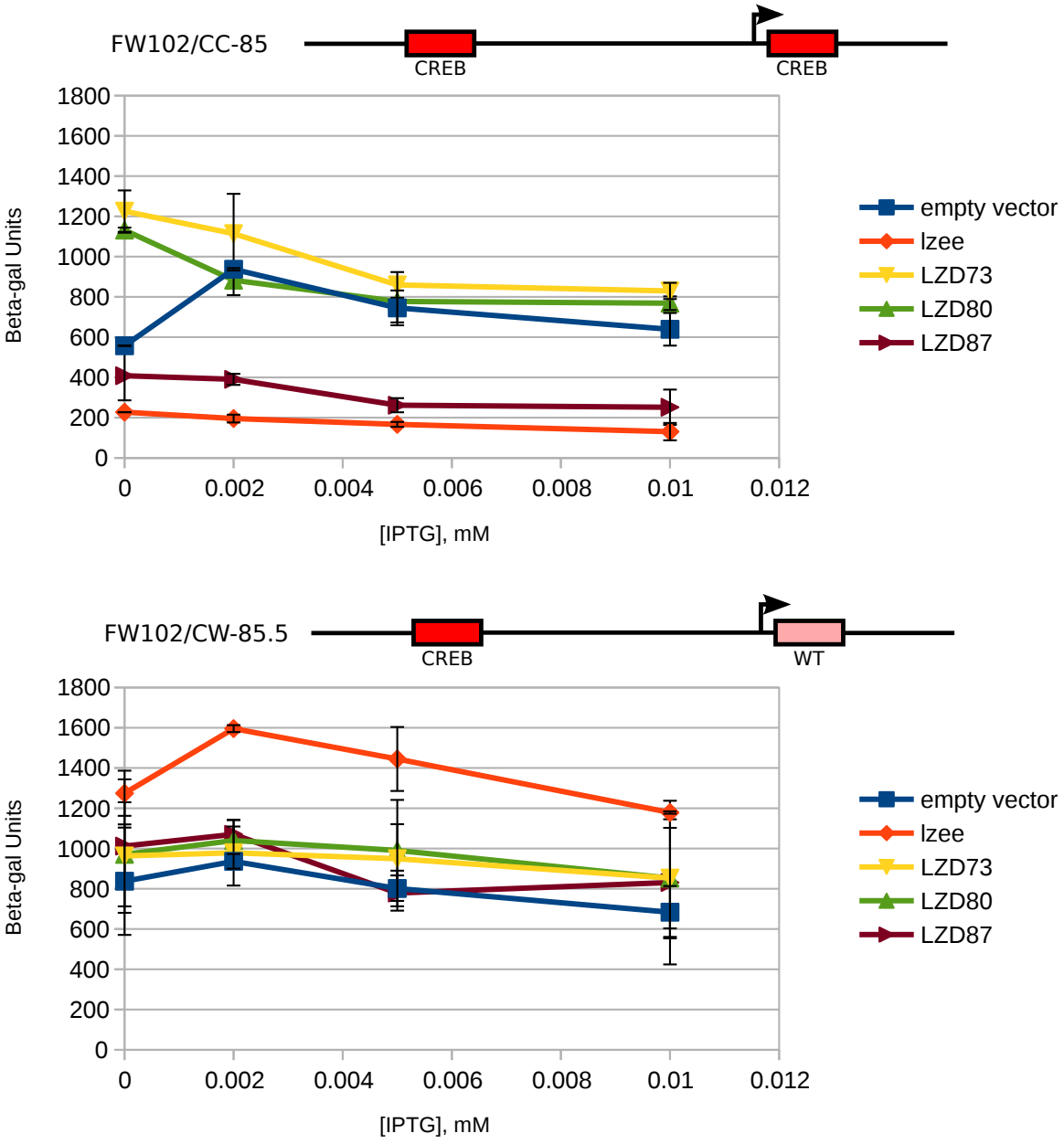


Figure 3.12: Comparing β-galactosidase activity in looping strains transformed with lzee and LZDs, showing the significantly lower β-galactosidase levels from lzee and LZD87 transformants containing the proximal CREB site. The values reported are average numbers from two replicates selected from LB X-gal plates

Table 3.4: Reporter protein behavior of lzee annd LZD proteins in looping strains. Results show weak repression of the lacZ protein from FW102/CC-85 strains upon induction of the recombinant proteins, but not the empty vector (tac-pGEX). On the other hand, there is no difference in lacZ protein expression from FW102/CW-85.5 strains upon recombinant protein expression.

Reporter Strain	Expression Plasmid	β-gal Units (0 mM IPTG)	β-gal Units (10 μM IPTG)	Repression Ratio, RR
FW102/CC-85	tac-pGEX	557 ± 1	639 ± 81	0.9 ± 0.1
FW102/CC-85	tac-lzee	227.3 ± 0.1	131 ± 43	1.7 ± 0.6
FW102/CC-85	tac-LZD73	1226 ± 103	829 ± 40	1.5 ± 0.1
FW102/CC-85	tac-LZD80	1132 ± 13	768 ± 34	1.47 ± 0.07
FW102/CC-85	tac-LZD87	408 ± 122	252 ± 88	1.6 ± 0.7
FW102/CW-85.5	tac-pGEX	837 ± 267	684 ± 130	1.2 ± 0.5
FW102/CW-85.5	tac-lzee	1275 ± 112	1179 ± 6	1.08 ± 0.10
FW102/CW-85.5	tac-LZD73	963 ± 156	853 ± 250	1.1 ± 0.4
FW102/CW-85.5	tac-LZD80	971 ± 259	853 ± 292	1.1 ± 0.5
FW102/CW-85.5	tac-LZD87	1012 ± 332	831 ± 407	1.2 ± 0.7

Figure 3.13 and Table 3.5 has shown some dose-dependent repression of the reporter protein in all of the transformants analyzed. Further reporter protein assays on BL876 strains transformed with lzee at 0 – 10 mM IPTG shows that β-galactoside repression indeed occurs in a dose-dependent manner (see Figure 3.14). There are two possible reasons for this observation. One would be due to leaky transcription from the *tac* promoter. Previous *in vitro* gel shift assays have shown that at excess amounts of protein, lzee starts to bind non-specifically to DNA. Another possibility is due to transcriptional load. Upon IPTG induction, there is a reduced number of RNA polymerase transcribing the reporter gene because some of the former are being used to transcribe the genes from the multi-copy plasmids (Glick, 1995).

Given the narrower range of IPTG concentrations required to do the reporter

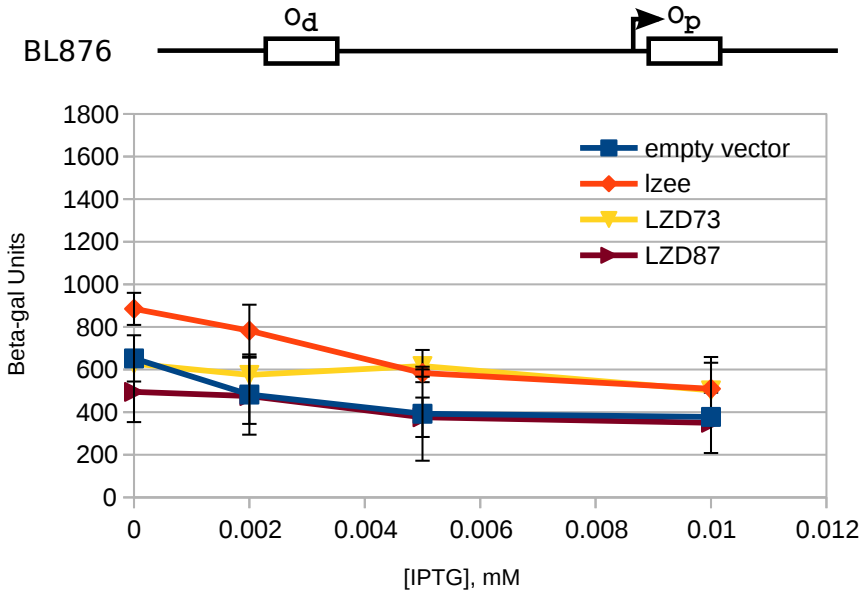


Figure 3.13: β-galactosidase activity in BL876 strains transformed with lzee and LZDs, showing dose-dependent repression from a reporter strain that do not contain any GCN4 sites. The values reported are average numbers from two replicates selected from LB X-gal plates

Table 3.5: Reporter protein behavior of lzee annd LZD proteins in BL876 strains, showing dose-dependent repression from a reporter strain that do not contain any GCN4 sites.

Reporter Strain	Expression Plasmid	β-gal Units (0 mM IPTG)	β-gal Units (10 μM IPTG)	Repression Ratio, RR
BL876	tac-pGEX	652 $\pm$ 109	377 $\pm$ 307	2 $\pm$ 1
BL876	tac-lzee	885 $\pm$ 75	509 $\pm$ 122	1.7 $\pm$ 0.4
BL876	tac-LZD73	627 $\pm$ 14	502 $\pm$ 157	1.2 $\pm$ 0.4
BL876	tac-LZD87	495 $\pm$ 142	350 $\pm$ 142	1.4 $\pm$ 0.7

assays without killing the host cells, it is reasonable to say that the LZD proteins are likely better repressors than lzee. The toxicity is most likely coming from non-specific binding of LZD proteins at multiple sites in the bacterial chromosome, inhibiting cell growth. This can be circumvented by cloning these proteins into

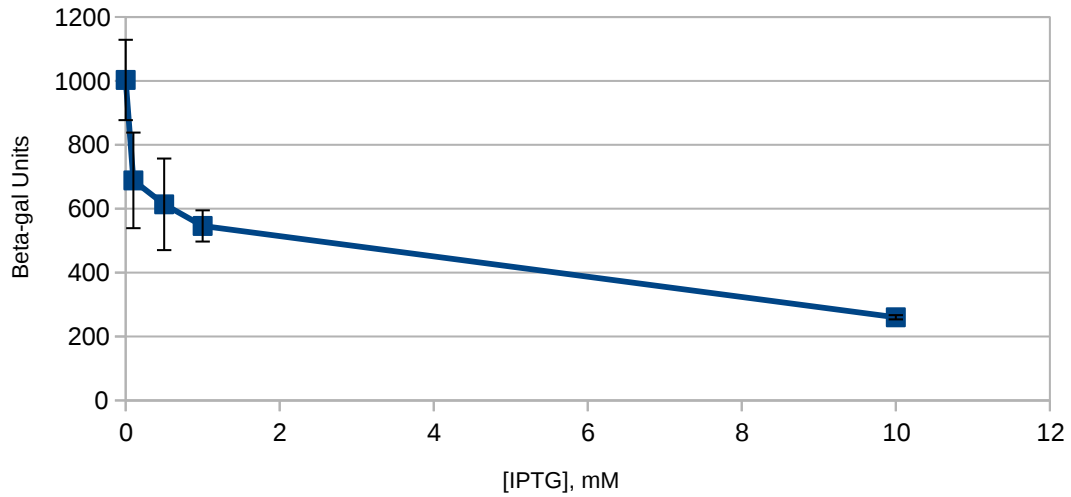


Figure 3.14: β-galactosidase activity of BL876 reporter strain, transformed with tac-lzee. Overexpression of lzee (using 10 mM IPTG) showed significant repression of the reporter protein.

an expression plasmid that minimizes, if not prevents, leaky transcription under non-inducing conditions.

3.4.4 Colorimetric screen on tac-pGEX transformants using MacConkey and tetrazolium media

The colorimetric screen results with Bluo-gal (or X-gal) indicate that white/ blue screening is not sensitive enough to distinguish repression by strong vs. weak binding at the primary operator. Thus, we used two other indicator media to characterize the behavior of our DNA-binding proteins. First, the non-looping reporter strains were inoculated into liquid cultures and plated onto both MacConkey lactose and tetrazolium lactose plates. XL10-Gold was used as the control for a lac$^-$ strain.

By design, XL10-Gold contains an incomplete fragment of lacZ (lacZΔM15) in its F' episome. The alpha fragment has to be provided *in trans*, usually from a multicopy plasmid, in order to generate a fully functional β-galactosidase.

All non-looping strains grew red colonies on MacConkey media, and white colonies on tetrazolium lactose (TZ lac) media, indicating that these strains are lac^+. On the other hand, XL10-Gold, which are lac^-, yielded white colonies on MacConkey plates, and red ones on TZ lac plates. Interestingly, when the strains were plated onto the indicator plates without antibiotics, the reporter strains retained the same colors, whereas XL10-Gold switched colors (from white to red in MacConkey, and from red to white in TZ lac). Producing antibiotic resistance is taxing to the host cell, as it uses up a significant portion of the host cell's resources [Bentley et al., 1990]. When *E. coli* cells containing customized F' episomes are grown without antibiotics, normal *E. coli* cells, which contain a wild-type lac operon in their chromosomes, would eventually outgrow the ones with antibiotic resistance. Thus, in losing the F' episome, there is loss of lac^- strains and gradual growth of lac^+ strains.

Transformation of the tac-pGEX6P1 recombinants into the nonlooping reporter strains generated pink to red colonies on MacConkey plates, and pale pink colonies on TZ lac plates. However, looping strains transformed with lzee or any of the LZD expression plasmids either (1) did not grow colonies on MacConkey and tetrazolium plates, as in the case of lzee and LZD87, or (2) surviving colonies have roughly 50% chance of surviving the liquid culture growth. In a control experiment, the same reporter strains were transformed with the LZD87-pRSETA plasmid. Expression of the gene inserts in pRSETA is regulated by the T7 promoter. Our

reporter strains do not produce T7 polymerase, thus there would not be any leaky transcription from the expression plasmid in the resulting transformants. Indeed, all LZD87-pRSETA transformants grew a lawn of white colonies on TZ lac plates, and light blue colonies on X-gal plates. This illustrates that over-expression of the DNA-binding proteins from pGEX tends to be detrimental to the host cells, which makes it quite difficult to characterize them using our *in vivo* genetic assays.

The pGEX6P1 plasmids that we are using as expression plasmids in our initial β-galactosidase assays are designed to produce high levels of recombinant protein using the minimum amount of inducer. While this may work for large-scale production of recombinant protein for *in vitro* experiments, this is not applicable for our *in vivo* assays, where recombinant protein levels have to be fine-tuned at below-saturating conditions. Therefore, another expression plasmid have to be used that would produce much lower quantities of our putative proteins that are more appropriate to *in vivo* repression assays.

3.4.5 Colorimetric screen on lacUV5-pGEX transformants

At this point, neither indicator media can elicit different phenotypes in distinguishing strong vs. weak repression without killing the host cells. However, tetrazolium-containing media is useful for detecting reporter strains with reduced *lac* activity, since strongly Lac^+ strains produce white colonies while weakly Lac^+ or Lac^- transformants produce pink or red colonies (Shuman & Silhavy, 2003). In addition, the indicator media recipe is easily adjustable to prepare inducible and

non-inducible plates by simply changing the sugar supplement. On the other hand, Xgal-containing media are still useful to detect transformants with defective promoters, as these transformants would form white colonies. Hence, we would be using Xgal and tetrazolium in our indicator media for our qualitative colorimetric screens.

In order to address the toxicity issue, the *tac* promoters in the expression plasmids were mutated to the *lacUV5* or wild-type *lac* promoters. Recombinant protein expression from the *lacUV5*-controlled plasmids were then verified in BL21 *E. coli* cells by Western blot that detects the GST affinity tag (see Figure 3.15).

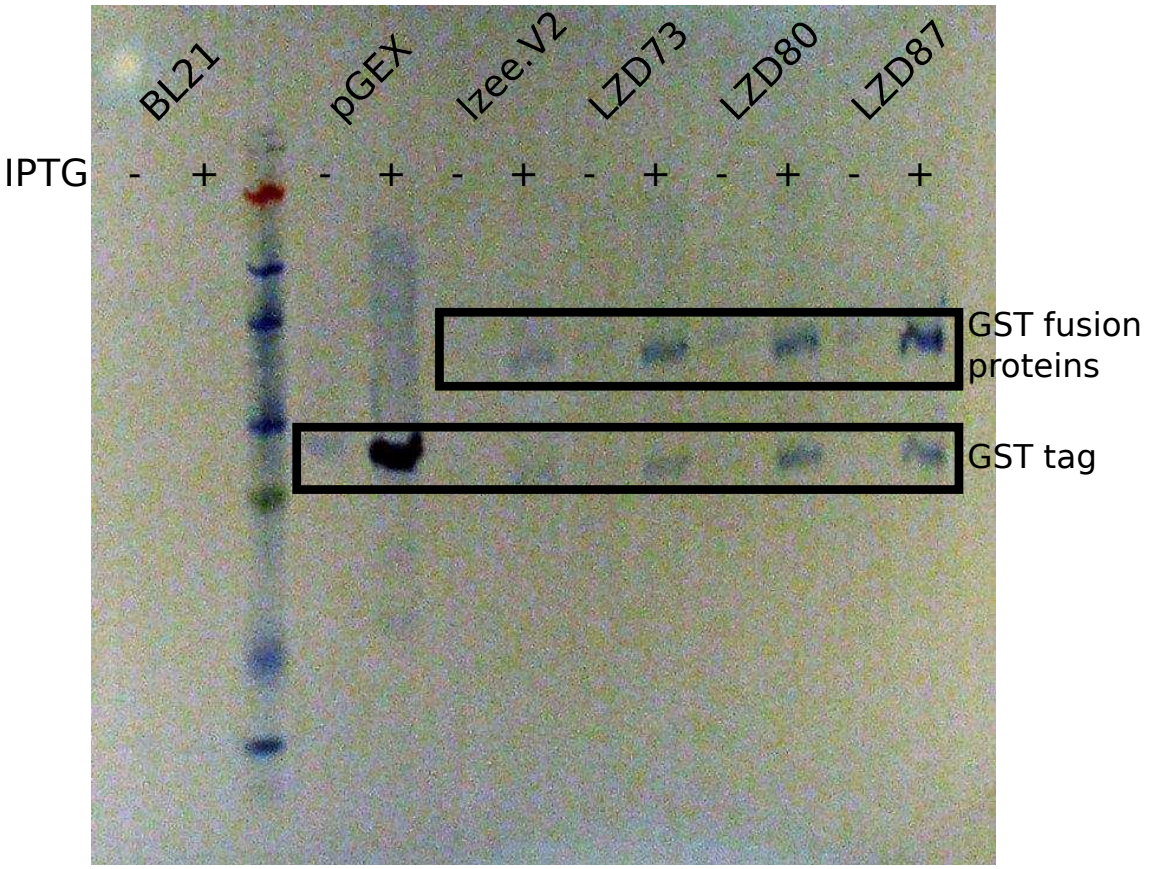


Figure 3.15: Western blot of clarified lysates from lacUV5-pGEX transformants (empty vector, lzeeV2, LZD73, LZD80, LZD87) in BL21 *E. coli* cells, 2 hours post-induction with 1 mM IPTG.

The lacUV5-pGEX recombinants (empty vector, lzeeV2 and LZD87) were transformed into FW102/CC-85 and plated onto X-gal and TZ lac plates. Lzee.v2 has the same protein sequence as lzee, but contains a few silent mutations, the rationale of which is explained in Chapter 5. All transformants grew colonies on both indicator plates, signifying that leaky transcription of proteins from the *lacUV5* promoter is not detrimental to host cell growth. The resulting blue colonies from X-gal plates, and white (empty vector) or light pink (lzee.v2 and LZD87) colonies from TZ lac plates were then analyzed for reporter protein activity. Results are shown in Figure 3.16. Both lzee.v2 and LZD87 transformants selected from X-gal plates showed a dose-dependent repression of β-galactosidase upon IPTG induction. Surprisingly, transformants selected from TZ lac plates showed significantly lower amounts of the reporter protein in our assays, even in non-inducing conditions. Transformants containing the empty vector selected from both plates showed comparable reporter activity, and do not seem to be affected upon IPTG induction.

The experiment was repeated on non-looping strains (FW102/proxCREB-1 and FW102/proxWT-1) and a longer looping strain (FW102/CW-237.5) to see if this observation only occurs when growing transformants containing the GCN4 binding protein on TZ lac media (see Figures 3.17 – 3.19). Reporter activity assays showed that there is a correlation between growing transformants on TZ lac and lowered amounts of β-galactosidase. The same observation is also seen when the reporter strains are transformed with the empty lacUV5-pGEX expression plasmid.

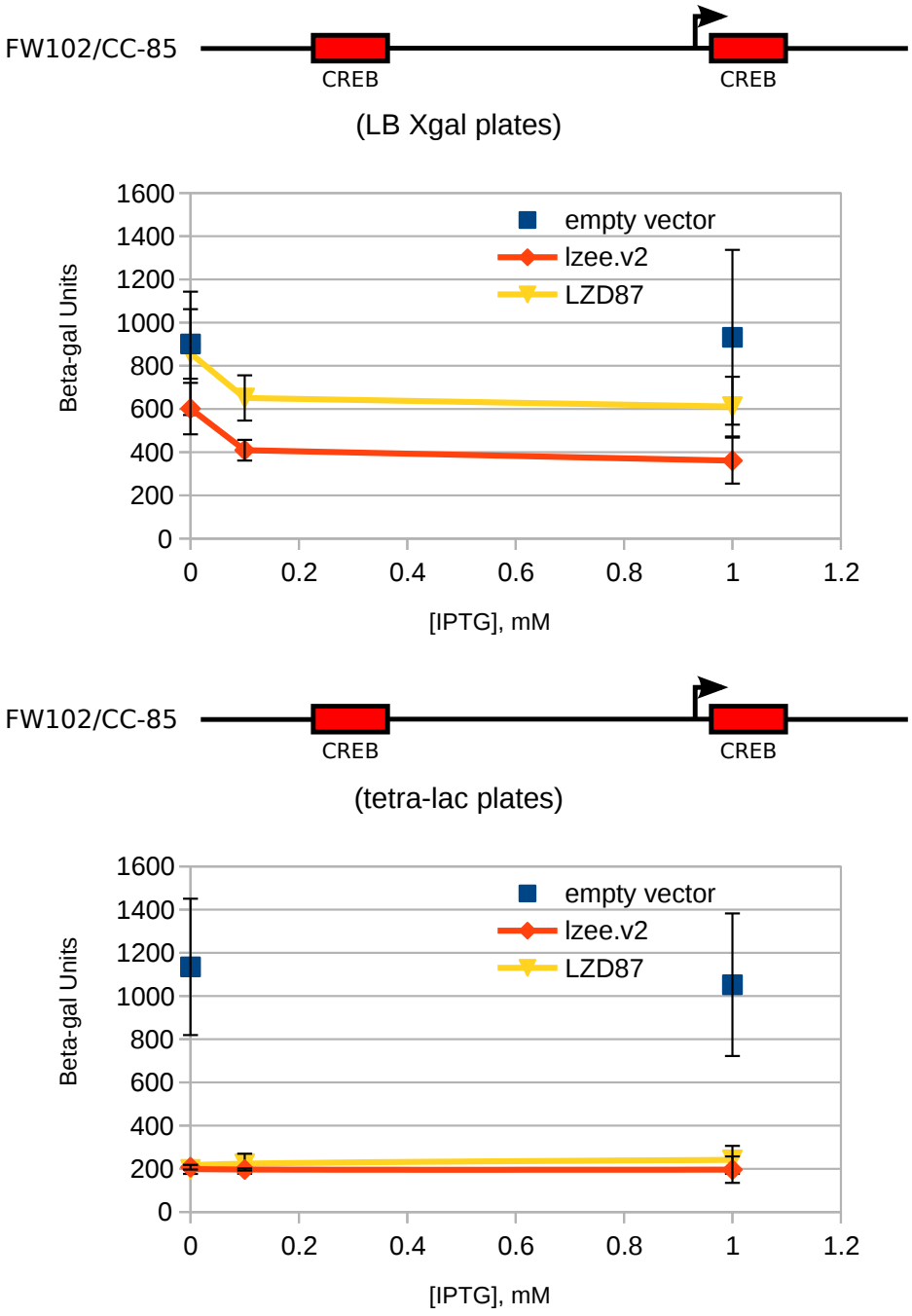


Figure 3.16: Comparing β-galactosidase amounts on FW102/CC-85 transformants selected from different indicator plates. Lzee.v2 and LZD87 transformants selected from Xgal plates showed dose-dependent repression of the reporter protein (top graph), while the same transformants selected from TZ lac plates exhibited very low levels of the reporter protein (bottom graph).

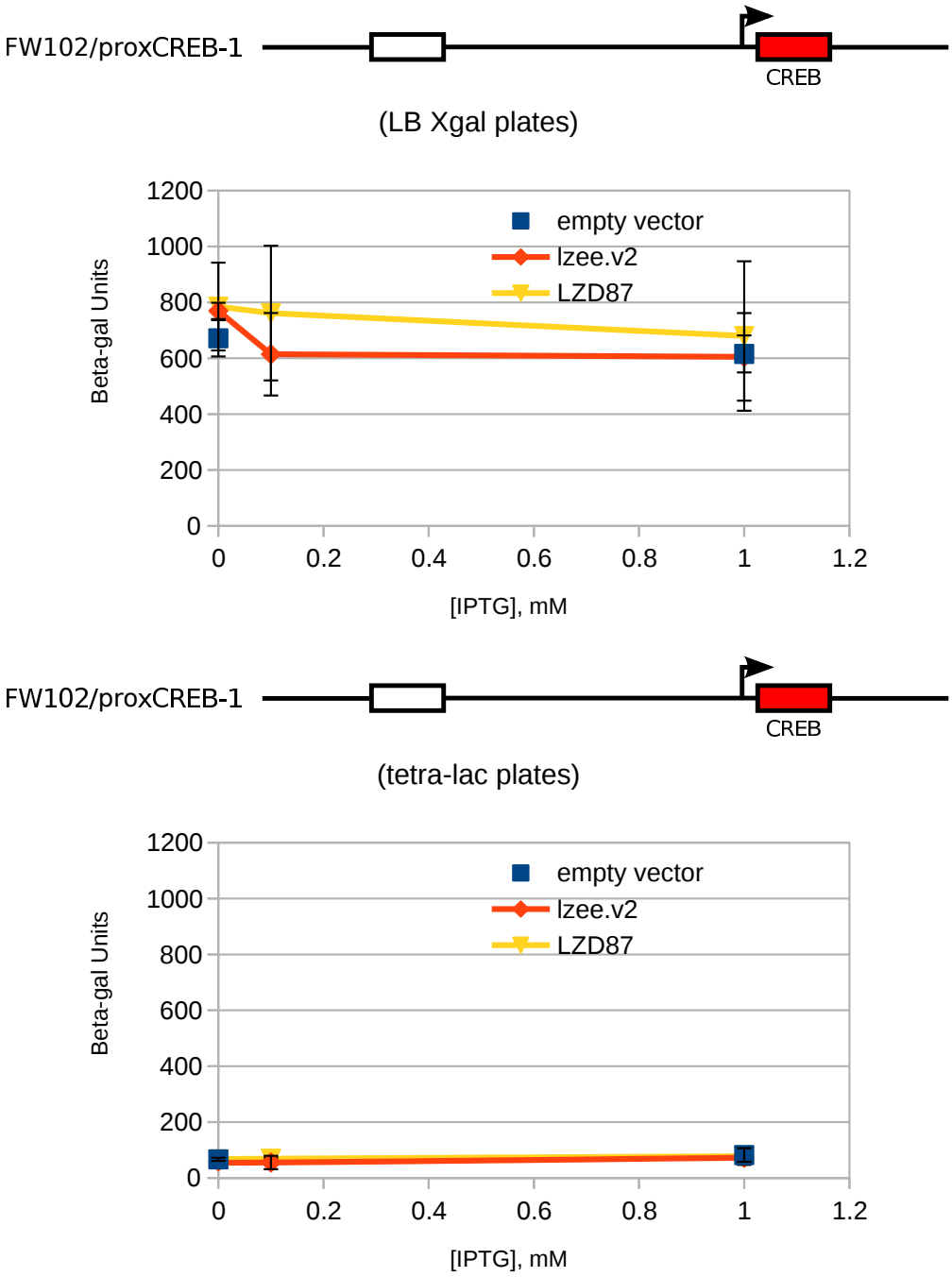


Figure 3.17: Comparing β-galactosidase amounts on FW102/proxCREB-1 transformants selected from different indicator plates. Transformants selected from Xgal plates showed no dose-dependent repression of the reporter protein (top graph), while the those selected from TZ lac plates exhibited significantly lower amounts of the reporter protein (bottom graph).

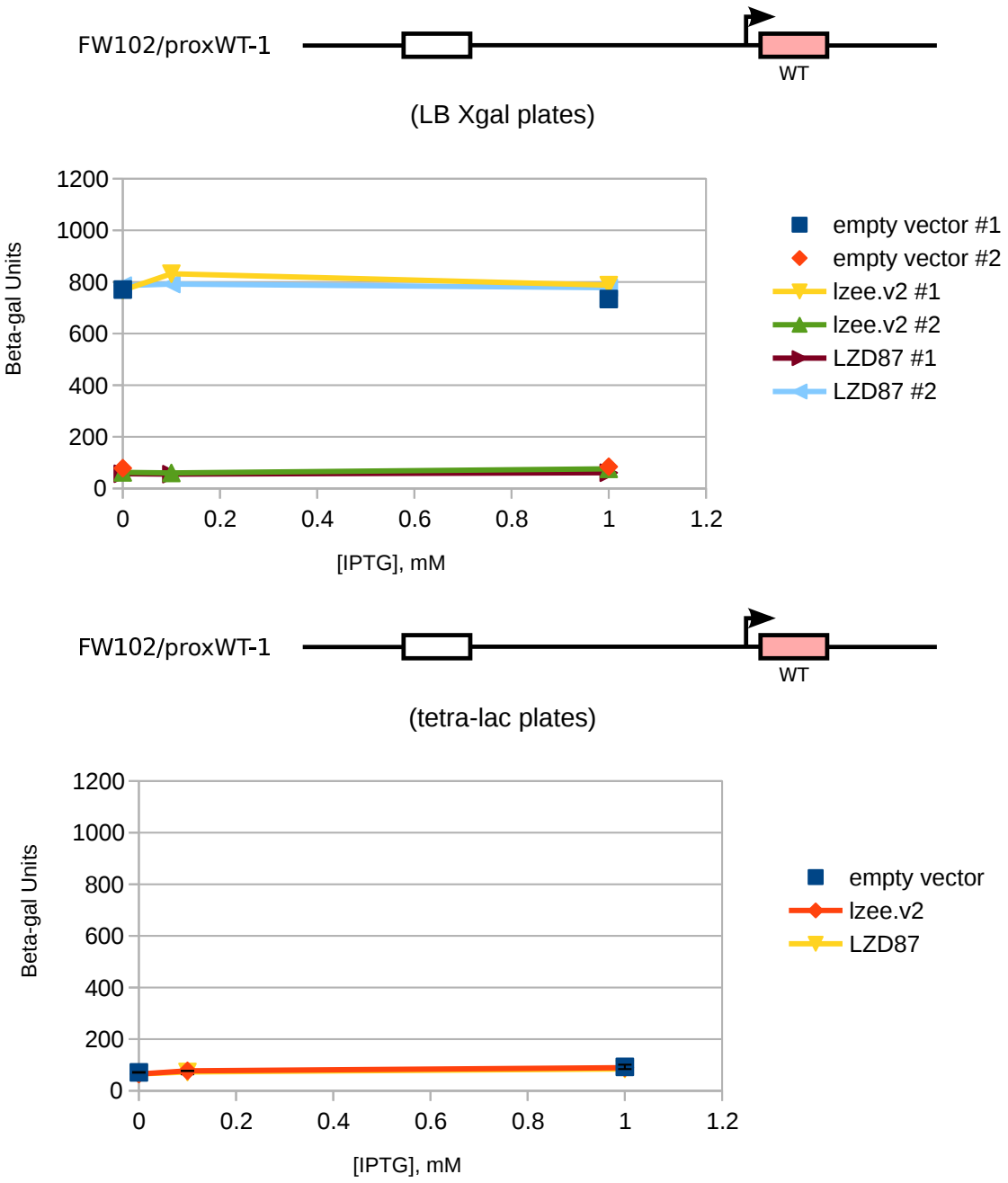


Figure 3.18: Comparing β-galactosidase amounts on FW102/proxWT-1 transformants selected from different indicator plates. Transformants selected from Xgal plates showed no dose-dependent repression of the reporter protein (top graph), while the those selected from tetrazolium-lactose plates exhibited significantly lower amounts of the reporter protein (bottom graph).

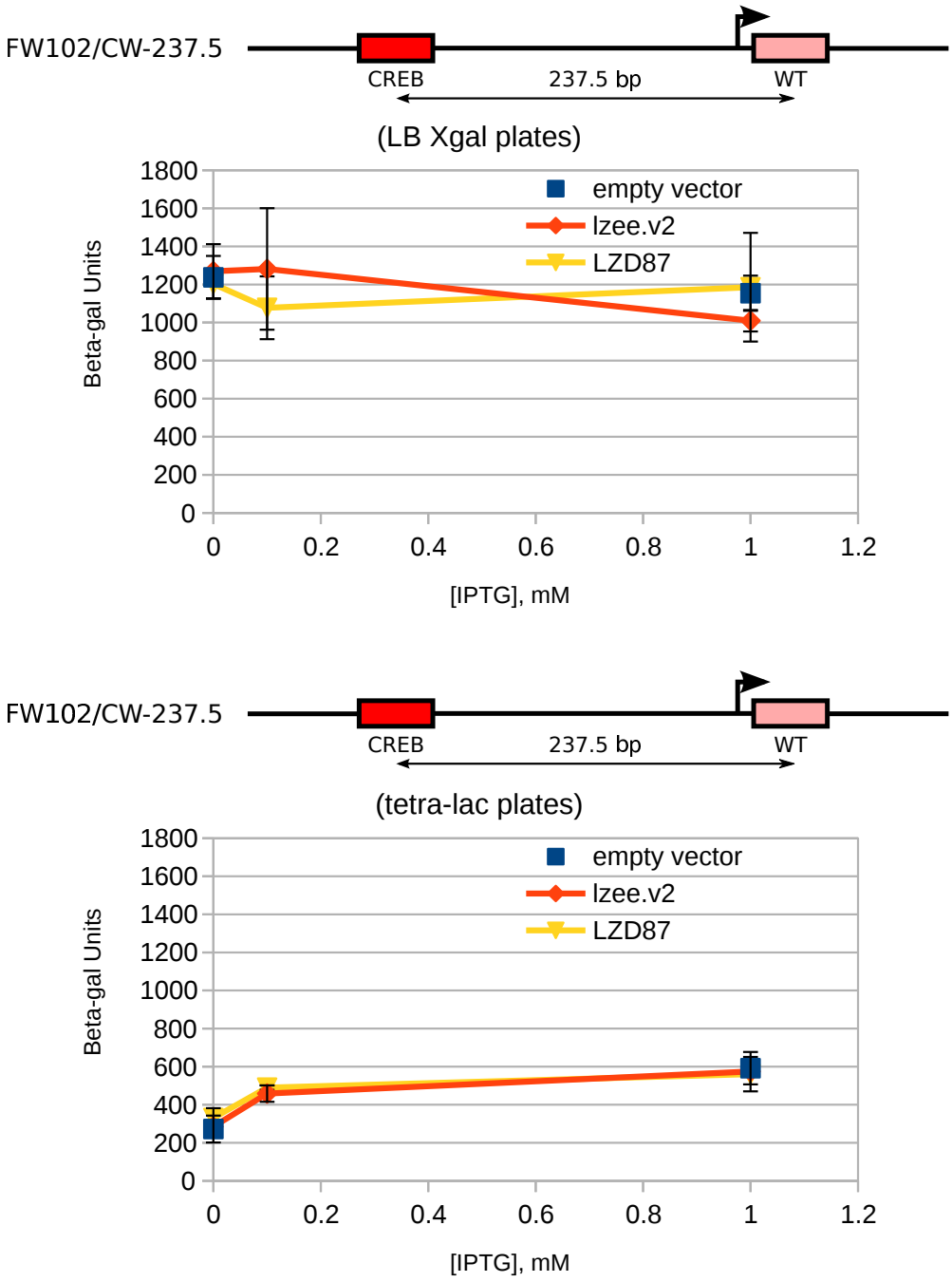


Figure 3.19: Comparing β-galactosidase amounts on FW102/CW-237.5 transformants selected from different indicator plates. Transformants selected from Xgal plates showed no dose-dependent repression of the reporter protein (top graph), while the those selected from TZ-lac plates exhibited significantly lower amounts of the reporter protein that increase slightly upon IPTG induction (bottom graph).

3.4.6 Cross-streaking experiments between indicator media

A series of cross-streaking experiments between X-gal and tetrazolium was subsequently done to find out (1) if the low levels of lacZ in the transformants from tetrazolium-lactose plates are triggered by the presence of the inducer in the plates, and (2) if this repression persists when the cells are moved to a new growth media that do not contain the inducer. It is possible that the low expression of the reporter protein could be due to the loss of the F' episome in the host cells. If this were indeed the case, we want to be able to pick out transformants that have intact F' episomes prior to doing the reporter protein assays. The looping strains FW102/CC-85, FW102/CW-85.5 and FW102/CW-237.5 were transformed with lacUV5-pGEX6P1 plasmids (empty vector, lzeeV2 and LZD87) and plated onto TZ glu, TZ lac, LB X-gal and LB X-gal IPTG indicator plates. Figure 3.20 displays the transformation results on four different indicator plates.

Most of the transformants yielded pink colonies on the tetrazolium plates, and similar amounts of blue and white colonies on X-gal ± IPTG plates. Transformants containing the empty vector grew comparable numbers of colonies on both tetrazolium and X-gal ± IPTG plates. Lzee.v2 transformants grown on X-gal plates generally have more colonies compared to the ones on X-gal IPTG plates. However, the same transformants did not grow colonies on TZ lac plates, but yielded a lot of pink colonies on TZ glu plates. The only exception was the FW102/CW-237.5 transformant, where there was no colony growth in TZ glu plates. LZD87 transformants have significantly fewer colonies on TZ glu plates compared to those with

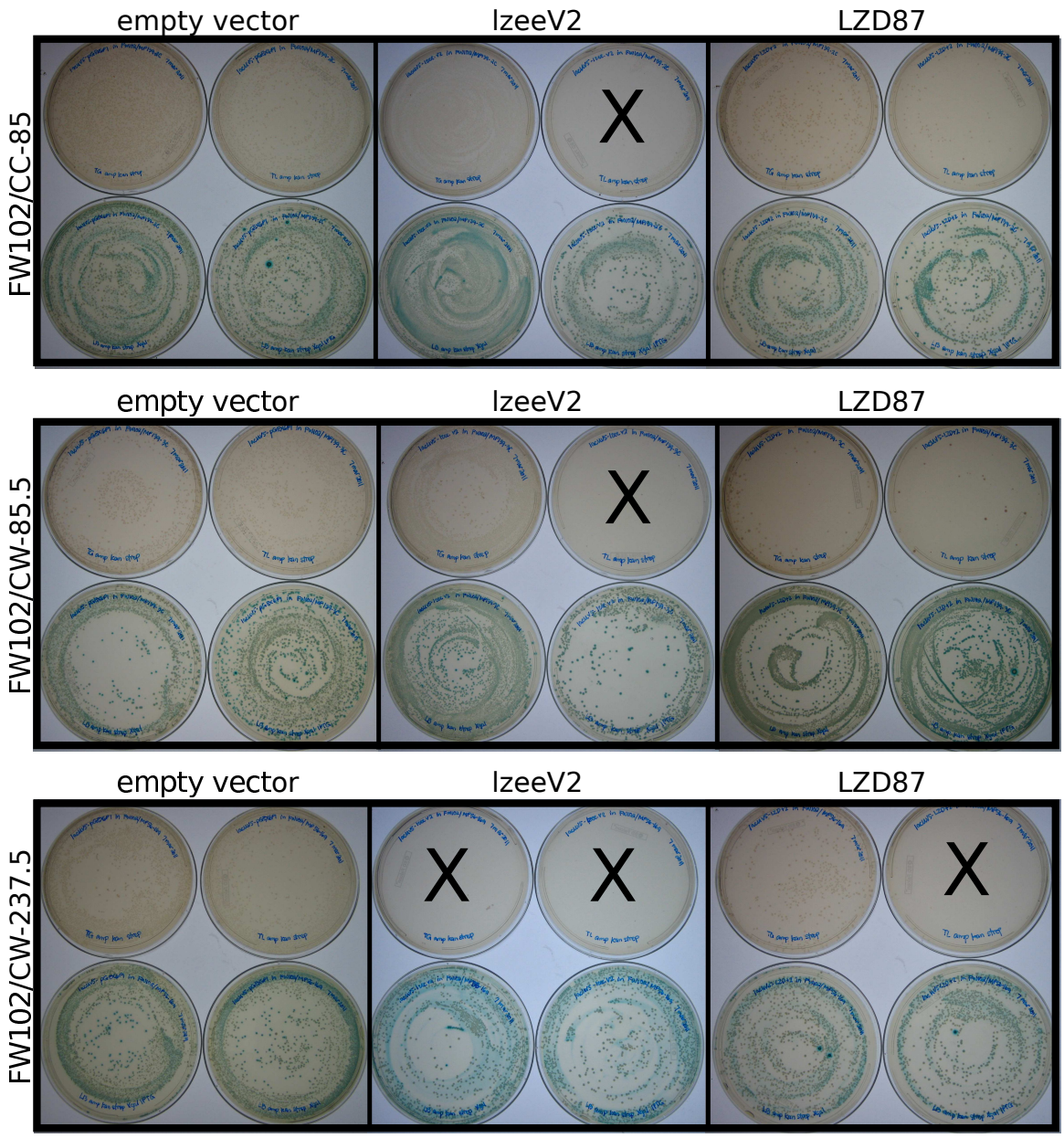


Figure 3.20: Colorimetric screens on lacUV5-pGEX transformants using tetrazolium and Xgal plates. Each square that corresponds to an expression plasmid:reporter strain transformant, contains the four selection plates. top row: TZ glu (left), TZ lac (right); bottom row: LB Xgal (left), LB Xgal IPTG (right). Plates marked with X indicate no colony growth. LZD87:FW102/CW-85.5 transformants grew sparse red colonies on both tetrazolium plates

the empty vector or lzee.v2. LZD87 transformed into FW102/CW-85.5 cells grew very few red colonies on tetrazolium-containing plates. The number of colonies are comparable upon IPTG induction (i.e. LB ± IPTG), but decreased significantly in the presence of lactose (tetrazolium ± lactose). Moreover, FW102/CW-237.5 cells transformed with LZD87 also did not grow colonies on tetrazolium-lactose plates. The following set of cross-plating experiments were then done using the colonies from these plates.

Initial plating results show that lzee.v2 becomes toxic to any reporter strain when grown in media containing lactose. To find out if the toxicity is triggered when the protein expression is induced on a selection plate, lzee.v2 transformants from Xgal plates were streaked onto TZ lac plates, and transformants from TZ glu plates were streaked onto Xgal IPTG plates. Results showed either no colony growth (FW102/CC-85 transformants) or red colonies on TZ lac plates, and distinct white colonies on Xgal IPTG plates. This shows that plating lzee.v2 transformants onto selection plates containing an inducer either renders toxicity to the host cells, or generates transformants with altered characteristics. Thus, subsequent repression assays on lzee.v2 were done on transformants grown on non-inducible selection plates.

Dark and light blue colonies from X-gal plates were streaked onto fresh TZ glu and TZ lac plates. All of the dark blue colonies resulted in pink colonies on TZ glu plates, and the light blue ones to red colonies. Most of the dark blue colonies from X-gal plates turned to pink colonies on TZ lac plates, and the light blue ones became red colonies.

Pink colonies from TZ glu plates were streaked onto fresh X-gal IPTG plates. FW102/CW-237.5 transformants selected containing the empty vector, and lzee.v2 grew blue colonies on the new selection plates. All of the LZD87 transformants left blue streaks with mostly white colonies.

Finally, blue and light blue colonies from X-gal IPTG plates were streaked onto TZ glu plates. Transformants containing the empty vector produced pink colonies from blue ones selected, and red colonies from light blue ones. On the other hand, all transformants containing the recombinant pGEX plasmid (lzee.v2, LZD87) generated red colonies, regardless of the color of the selected colony from X-gal IPTG.

A few points can be deduced from the results presented. First, colony color on uninduced indicator plates can be used to qualitatively assess the behavior of the selected transformant in the reporter protein assays. Cross-plating results from LB Xgal to TZ glu plates should just reiterate the phenotype from LB Xgal since TZ glu should be similar as LB Xgal. Hence, light blue to white colonies would yield red colonies, indicative of a lac^- phenotype. We would then expect these transformants to show lower amounts of the reporter protein at non-inducing conditions.

Second, cross-plating from uninduced indicator plates to indicator plates containing IPTG or lactose did not show a clear change of phenotype, i.e. blue colonies from X-gal plates generating red colonies in TZ-lac plates. However, expression of lzee transformants by lactose induction, either renders toxicity to the host cell, or generates survivors with altered phenotypes.

Finally, cross-plating of transformants from selection plates containing the

inducer to indicator plates that do not contain the inducer shows that any transformants with altered characteristics have maintained that phenotype upon removal of the inducer. This introduces the possibility of reporter protein repression by our DNA binding proteins occurring via a different mechanism.

3.4.7 Repression assays on 85- to 650-bp looping strains

In order to minimize transcriptional load, one approach was to mutate the *tac* promoters in all our recombinant plasmids into either *lacUV5* or wild-type *lac* promoters. In addition, longer looping strains were constructed to compare looping efficiencies *in vivo* over a wider range of distances between GCN4 binding sites. The recombinant expression plasmids were transformed into the various looping and non-looping reporter strains to compare the extent of reporter protein repression over the span of 85 – 650-bp distance between GCN4 sites. Table 3.6 summarizes the repression ratios of the recombinant proteins in reporter strains containing the CREB site in the downstream position.

Table 3.6: Repression ratios on reporter strains containing proximal CREB site. All transformants analyzed were selected from LB plates. Results from BL876 transformants are included to provide baseline repression ratio values

Reporter Strain	no plasmid	pGEX	lzee.v2	LZD73	LZD80	LZD87
BL876	1.04 ± 0.06	1.21 ± 0.05	1.37 ± 0.01	1.26 ± 0.07	1.11 ± 0.06	1.15 ± 0.01
proxCREB-1	1.01 ± 0.06	1.22 ± 0.05	2.39 ± 0.02	1.44 ± 0.04	1.53 ± 0.04	2.29 ± 0.02
CC-85	0.99 ± 0.03	not done	2.45 ± 0.02	1.43 ± 0.04	1.74 ± 0.06	2.482 ± 0.003
proxCREB-2	1.1 ± 0.3	1.29 ± 0.04	2.05 ± 0.06	1.2 ± 0.1	1.58 ± 0.06	2.00 ± 0.09
IC-450	1.1 ± 0.1	1.3 ± 0.1	2.31 ± 0.02	1.34 ± 0.02	1.58 ± 0.06	2.24 ± 0.04
IC-650	1.2 ± 0.2	1.25 ± 0.05	1.78 ± 0.05	1.5 ± 0.1	1.59 ± 0.03	1.66 ± 0.07

The first thing to note is the small error associated with most of the repression ratios calculated, which shows similar behavior among transformants analyzed at sub-saturating amounts of the DNA binding proteins. Next, lzee.v2 and LZD87 transformants have relatively higher repression ratios compared to LZD73 and LZD80 transformants. This is most likely from the fact that the DNA-binding domain in lzee has high affinity to the CREB site, and can also bind to one half of the INV2 site. On the other hand, the DNA-binding domains in the LZD proteins have different affinities to the CREB site. The N-terminal end has higher binding affinity to the CREB site, while the C-terminal end binds more preferrably to the INV2 site. However, we could not tell at this point which end of the protein is actually bound to which site in our repression assays.

There are two possible explanations for the differences in the extent of repression among the proteins tested. One, the relative orientations of the DNA-binding domains in LZD73 and LZD80 may not be optimal in stabilizing the DNA loops in the lengths between GCN4 sites analyzed. Another possibility could be the solubility issues for LZD73 and LZD80.

Comparison of the repression ratios between the looping and non-looping strains shows that there is slight enhancement of repression at the 85-bp distance (LZD80, LZD87), 450-bp distance (lzee.v2, LZD73, LZD87), and 650-bp distance (LZD73). We can not determine whether this increased repression is due to looping since there are not enough data to cover a distance range that spans one helical turn of the DNA.

Table 3.7 summarizes the reporter protein activities of host cells containing

the downstream wild-type GCN4 site and strong GCN4 sites from 138.5 to 450.5 bp away from the primary binding site, in the presence of lzee.v2 or LZD proteins. A comparison between the non-looping and the longer (300.5- and 450.5-bp) looping strains shows that there is no evidence of enhanced repression at this distance. This probably stems from the proximal site being too weak to keep the putative looping proteins bound long enough to repress the reporter gene.

Table 3.7: Repression ratios on reporter strains containing proximal wild-type GCN4 site. FW102/proxWT-1 and FW102/CW-237.5 transformants (last two rows) were selected from tetrazolium-lactose plates, the rest from LB plates. Results from BL876 transformants are included to provide baseline repression ratio values. No colonies – no colony growth in TZ-lac plates, Dead culture – selected colony from TZ-lac plate failed to grow in liquid growth medium, a – one replicate showed very low starting amounts of lacZ, b – both replicates showed very low starting amounts of lacZ.

Reporter Strain	no plasmid	pGEX	lzee.v2	LZD73	LZD80	LZD87
BL876	1.04 ± 0.06	1.21 ± 0.05	1.37 ± 0.01	1.26 ± 0.07	1.11 ± 0.06	1.15 ± 0.01
CW-138.5	1.1 ± 0.2	1 ± 2[a]	1 ± 2[a]	0.2 ± 3[b]	0.1 ± 1[b]	0.2 ± 0.7[b]
CW-187.5	1.1 ± 0.1	1.2 ± 0.1	1.3 ± 0.1	1 ± 1[a]	1.1 ± 0.2	0.1 ± 100[b]
proxWT-2	1.06 ± 0.04	1.21 ± 0.03	1.06 ± 0.03	1.19 ± 0.05	1.23 ± 0.04	1.15 ± 0.05
IW-300.5	1.0 ± 0.1	1.16 ± 0.07	1.09 ± 0.09	1.20 ± 0.06	1.24 ± 0.05	1.22 ± 0.06
IW-450.5	1.01 ± 0.07	1.2 ± 0.1	1.0 ± 0.1	1.19 ± 0.02	1.17 ± 0.09	1.09 ± 0.02
proxWT-1	1.0 ± 0.2	not done	dead culture	1.1 ± 0.1	1.1 ± 0.1	no colonies
CW-237.5	1.0 ± 0.1	not done	dead culture	1.153 ± 0.007	1.1 ± 0.1	no colonies

However, things get pretty interesting in the 138.5- to 237.5-bp distance range. Some of the transformants containing any of the pGEX expression plasmids exhibit signs of permanent repression in our liquid culture assays. The behavior seems to occur more frequently in (1) FW102/CW-138.5 transformants, where at least one replicate from each expression plasmid showed very low levels of lacZ, and in

(2) LZD73 and LZD87 transformants, where both of them gave aberrant average reporter protein activities at the 138.5- and 187.5-bp looping strains. A closer inspection on the reporter assay plots on LZDs:FW102/CW-138.5 transformants show that the reporter protein is actually activated upon IPTG induction (see Figure 3.21). Moreover, plating the transformants onto indicator plates that contain an inducer for the pGEX gene inserts prior to starter culture growth accelerates the process to the point that it kills the host cells. This is the case seen in lzee.v2 and LZD87 expression plasmids transformed into FW102/proxWT-1 and FW102/CW-237.5 reporter strains.

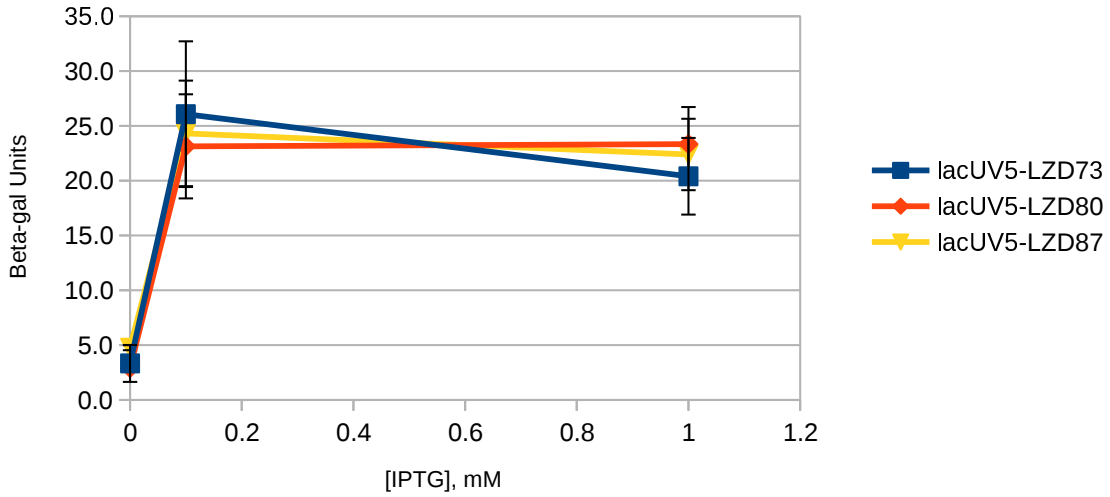


Figure 3.21: β-galactosidase levels in FW102/CW-138.5 reporter strains transformed with the lacUV5-LZDs, showing activation, instead of repression, of the reporter protein

3.4.8 Proposed model for baseline repression triggered by pGEX expression

The peculiar behavior seen in the shorter looping strains containing the pGEX recombinants suggests the following questions:

1. What causes complete repression of the reporter protein in the presence of the pGEX plasmids when plated on LB Xgal plates?

2. Why does this behavior occur less readily in reporter strains with longer separation between GCN4 sites?

To answer the first question, we have examined the sequences of the expression plasmids, and the *lacI* – *lacZ* cassette region in the F' episomes of all the reporter strains. A closer inspection of the complete DNA sequences of the pGEX plasmids revealed a wild-type *lac* promoter between the 3' end of $lacI^q$ gene and upstream of the *tac* (and *lacUV5*) promoter (see Appendix 2.1). This lac promoter has a lac O_1 operator and a short segment of the *lacZ* gene. This discovery has two possible implications: (1) leaky expression from this promoter could contribute to levels of the recombinant gene downstream of the main promoter, and (2) the *lacZ* gene sequence downstream of the wild-type promoter has high sequence homology with the *lacZ* sequence with the reporter gene in the F' episome. The first implication would explain the reduced lacZ levels observed in the reporter strains containing the proximal CREB site transformed with either tac-lzee or tac-LZD87 (see Figures 3.9, 3.10, and 3.12). The second one would have two possible consequences, as

summarized in Figures 3.22 and 3.23.

First, recombination at both *lacI* and *lacZ* regions introduces all three *lac* operators into the F' episome, and transfer the kanamycin resistance gene into pGEX. The F' episome would then lose the antibiotic resistance, and may be lost from the host cells. Hence, blue-white screening with Xgal would render the transformants white because there is no fully functional β-galactosidase protein that was provided by the F' episome. However, white-red screening with TZ-lac plates would have two possible outcomes. First, lactose permease (lacY protein product) is required for the host cells to metabolize lactose. With the F' episome gone, the resulting transformants can not utilize the added sugar and may not grow on lactose-containing media. This is most likely the case with lzeeV2 and LZD87 transformants not growing on TZ-lac plates. Second, any surviving transformants, which may not need to metabolize lactose, would most likely grow as pink to red colonies. This was the observation shown by LZD87:FW102/CW-85.5 transformants (see Figure 3.20).

Second, recombination events occuring at both the secondary *lac* promoter and *lacZ* regions also introduces two *lac* operators into the F' episome, while maintaining kanamycin resistance. The lactose repressor, overexpressed from the pGEX plasmids, can bind to this operator, which can inhibit *lacZYA* expression in the F' episome at non-inducing conditions. The resulting transformants would be blue on Xgal plates, and grow white to pink colonies on TZ lac plates. This would most likely explain the permanent repression or very low levels of reporter protein seen in the FW102/CW-138.5 strains transformed with the LZD proteins (see Figure 3.21).

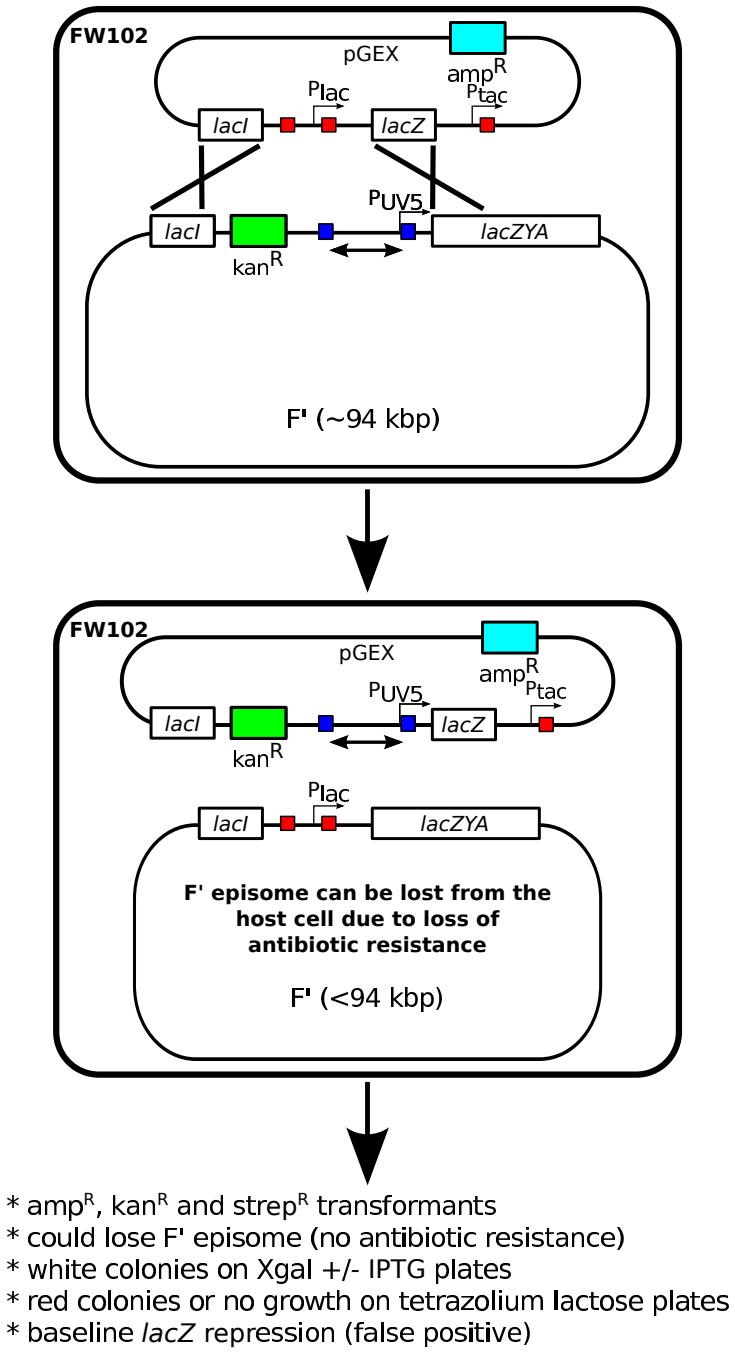


Figure 3.22: Proposed recombination event I in reporter strains transformed with pGEX plasmids. Double recombination events at the *lacI* and *lacZ* sequences introduces three *lac* operators (red squares; the third operator is within lacZ) into the F' episome, and transfers the kanamycin resistance gene into pGEX. The F' episome is kicked off the host cell, and the resulting transformants can not react with IPTG, or metabolize lactose. The double arrows indicate varying distances between the GCN4 sites. Blue squares – GCN4 site, red squares – lac O_1 and O_3 operators

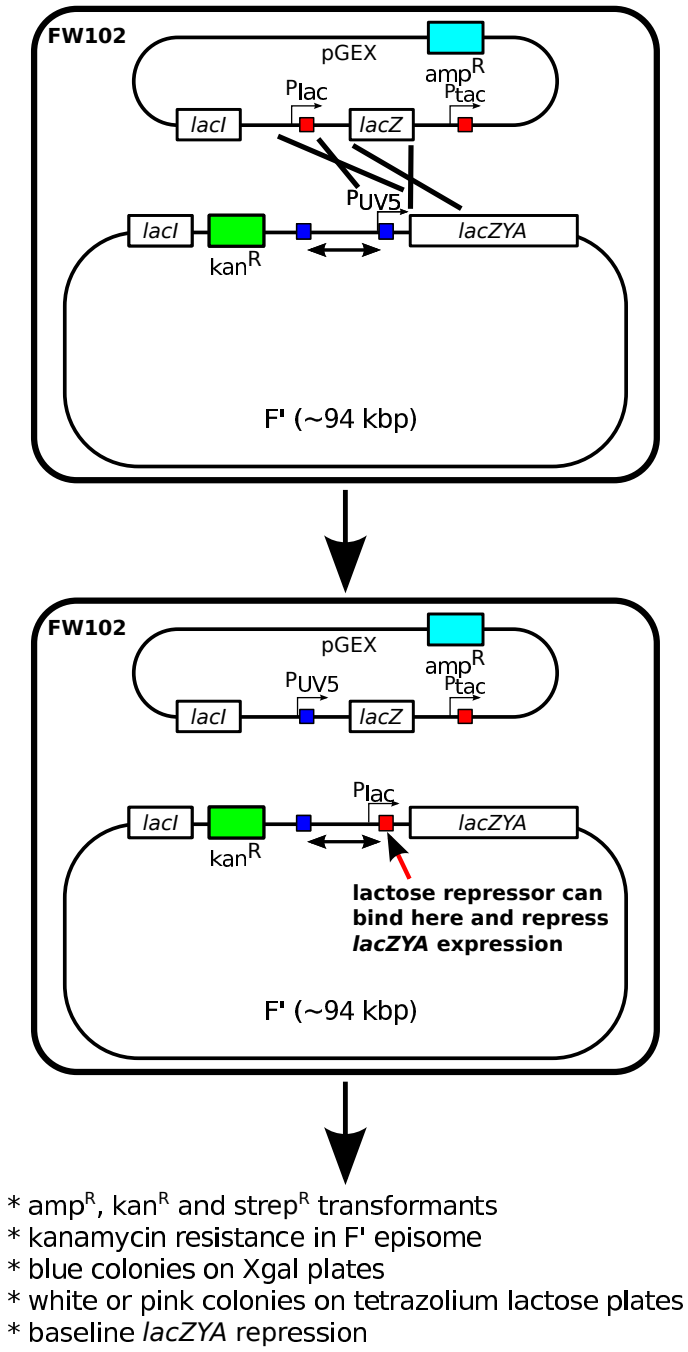


Figure 3.23: Proposed recombination event II in reporter strains transformed with pGEX plasmids. Double recombination events at the secondary *lac* promoter and *lacZ* sequences introduces two *lac* operators (red squares; the second operator is within lacZ) into the F' episome, while maintaining the kanamycin resistance gene. The resulting transformants can utilize IPTG and lactose, but will exhibit permanent repression in the liquid assays. The double arrows indicate varying distances between the GCN4 sites. Blue squares – GCN4 site, Red squares – lac O1 and O3 operators

Our colorimetric screen results have shown both events to be occurring randomly, although there is no clear trend as to which one is favored. In order to differentiate between the two, first thing to do is isolate the expression plasmids from the host cells post-induction. If *lacI* – *lacZ* recombination has occured, one telltale consequence would be a big difference in the expression plasmid size vs. the unrecombined plasmid size. Another experiment that could be done is to re-transform the plasmids into a different *E. coli* strain, and screen for ampicillin and kanamycin resistant transformants. Lastly, one can verify for the presence of the UV5 promoter – GCN4 operator sequence by PCR, or DNA sequencing.

We have routinely done colony and culture PCRs on the transformants we analyze in our genetic assays to verify that we are not losing either the expression plasmid nor the F' episome. All of the transformants we have analyzed so far have maintained both the plasmid and the F' episome. However, the presence of the correct PCR products from the colony or culture samples does not tell us where these products are being amplified from, although the presence of extraneous PCR products from transformants with altered phenotypes (i.e. white colonies from Xgal plates, or red colonies from tetrazolium plates) would support our recombination hypothesis. Analyzing the plasmids isolated from cultures post-pGEX expression, instead of just the cultures themselves for multiple PCR products would show irrefutable proof in support for our proposed explanation.

If the P_{lac} – *lacZ* recombination has happened, there would not be any difference in the plasmid sizes between the original and the modified plasmids. However, one can amplify the promoter between *lacI* and *lacZ* regions in the expression plas-

mids and determine the sequence. The proximal GCN4 site would be present in the recombined plasmids.

It is not clear from the proposed recombination events why they occur more frequently in shorter looping strains, or in the presence of the recombinant pGEX plasmids, but not the empty vector. Moreover, the recombination events could not fully explain why lzee.v2 transformants always end up dead when plated on TZ-lac plates, although we have not verified yet whether the cell toxicity is dependent on the presence of GCN4 sites in the reporter strains. One possibly farfetched idea would be a formation of a lzee-lactose repressor hybrid protein that brings the homologous regions from the F' episome and the expression plasmid closer together to initiate recombination. Based on what we have seen so far, there seems to be a link between induction of our proteins and increased recombination activity.

3.4.9 Reporter assays on lacUV5-pGEXb recombinants

The proposed recombination events due to sequence homologies between the F' episome and the expression plasmids complicate the interpretation of data from our looping assays. Thus, one way to minimize, if not prevent, recombination was to remove the secondary *lac* promoter – *lacZ* fusion in the recombinant pGEX plasmids. Recombination would most likely still happen at the *lacI* end, but this should not affect the results of the *in vivo* looping assay.

FW102/IC-450 reporter strains were transformed with the new pGEX recombinants (empty vector, lzee.v2b, LZD73b, LZD80b and LZD87b), and plated onto

TZ-glu and TZ-lac plates. All transformants grew mainly pink colonies on glucose-containing plates. On the other hand, transformants grown on TZ-lac plates are small light pink colonies, except for lzee.v2 transformants which did not grow at all. There was no significant difference in the number of colonies between the two colorimetric plates, indicating that lactose induction of the GST tag (from the empty vector) and LZDs is not detrimental to the host cell survival.

Next, we screened the different transformants for the presence of intact *lacUV5* promoter-GCN4 operator with Od-fp2.1 and Op-rp1.1. Culture PCR results from the rare red colonies selected from both plates, and white colonies from the uninduced plates showed the disappearance of the 1233-bp PCR product. In its stead were multiple PCR products, illustrating that these transformants may have undergone recombination, although we don't know exactly how. On the other hand, culture PCR results on pink colonies selected from both TZ-glu and TZ-lac plates showed the same single PCR product, indicating no recombination had occured from these transformants (see Figure 3.24). Thus, we proceeded on selecting pink colonies from the tetrazolium-containing plates in our next round of reporter assays.

In a new experiment, PCR on the pink colonies with recomFP2-lacZ + recomRP4-lacZ (which amplifies the suspected lacZ homologous region in the F' episome) did not generate a PCR product. In parallel, the gene insert in the expression plasmids are intact, albeit in relatively low amounts, as seen by PCR with seqFP2-pGEX + pGEXrp (see Figure 3.25).

It is possible that the recombination primers designed for amplifying the *lacZ* region in the F' episome are not optimal, so we proceeded to analyze the cultures

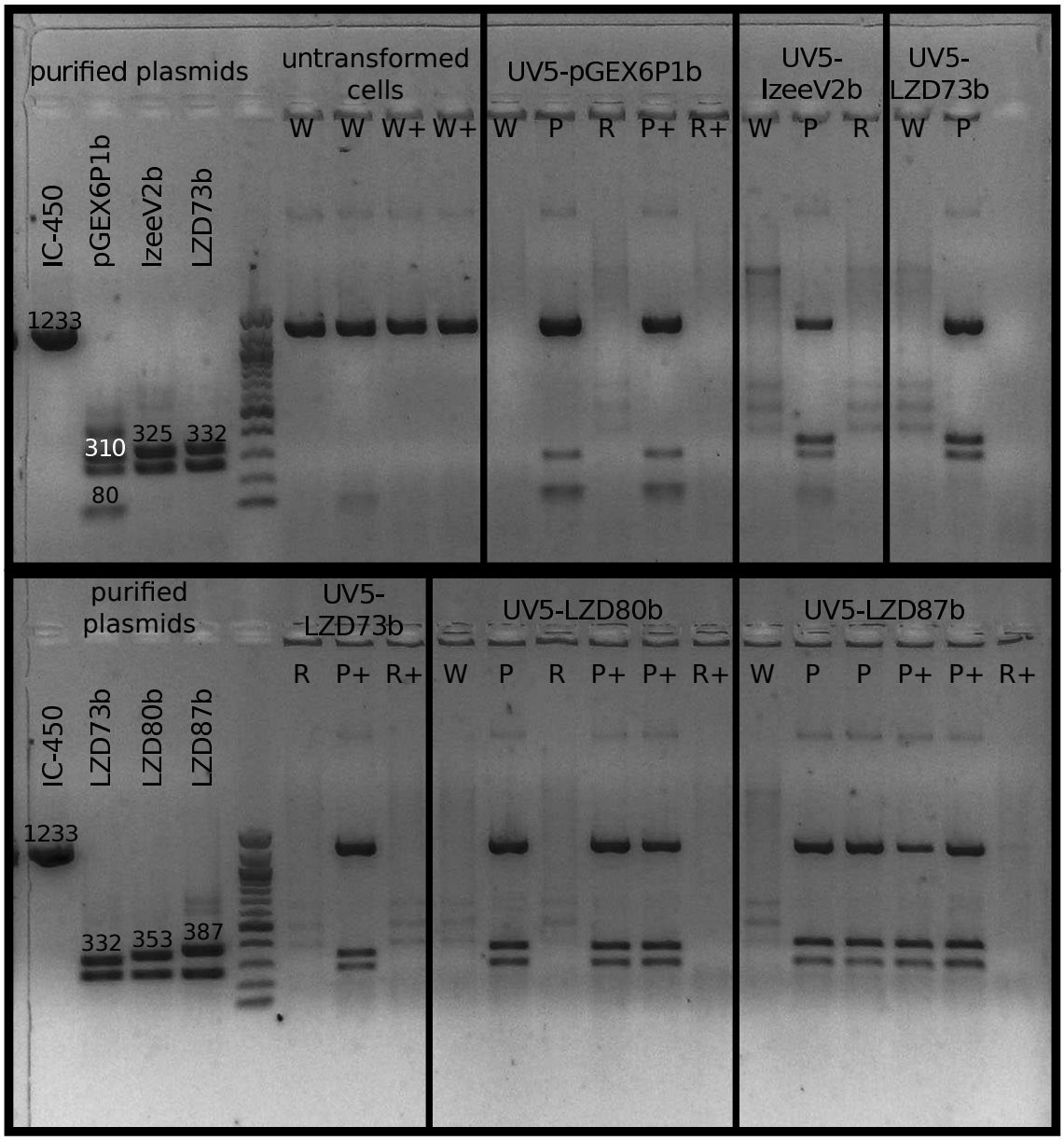


Figure 3.24: Culture PCR results I on FW102/IC-450 reporter strains transformed with lacUV5-pGEXb (empty vector, lzeeV2, LZD73, LZD80, LZD87), using Od-fp2.1 + Op-rp1.1 (1233 bp PCR product) and pGEXfp2 + pGEXrp primers (80, 325, 332, 353 or 387 bp PCR product). White (W, W+), light pink (P, P+), and red (R, R+) colonies were selected from tetrazolium-glucose and tetrazolium-lactose (+) plates for overnight growth in LB amp kan strep. Purified plasmids are used as controls for PCR. Extraneous PCR products (i.e. $\sim$ 250 bp and 310 bp) are due to PCR artifacts from one of the primers. Results show that the promoter-operator region in the F' episome has not been recombined into the pGEX expression plasmid, i.e. P and P+ samples give the same PCR products.

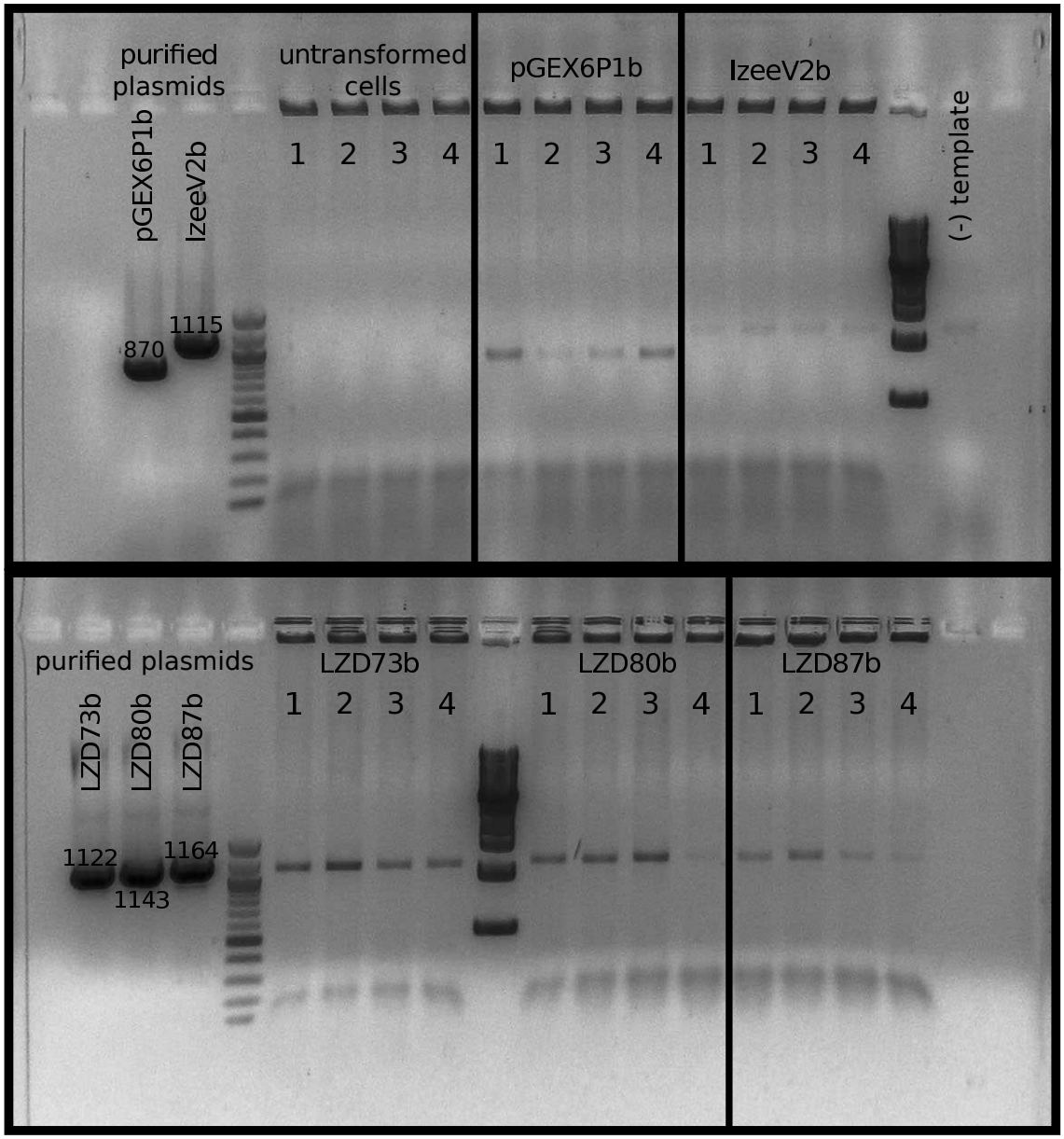


Figure 3.25: Culture PCR results II on FW102/IC-450 reporter strains transformed with lacUV5-pGEXb (empty vector, lzeeV2, LZD73, LZD80, LZD87), using recomFP2-lacZ + recomRP4-lacZ and seqFP2-pGEX + pGEXrp primers (870, 1115, 1122, 1143 or 1164 bp PCR products). Pink colonies were selected from TZ-glu (lanes 1 and 2) and TZ-lac plates (lanes 3 and 4) for overnight growth in LB amp kan strep. Purified plasmids are used as controls for PCR, shown on the left of the 100-bp DNA ladder. No products were detected in any of the transformants from recomFP2-lacZ + recomRP4-lacZ, which may indicate these primers to have very low annealing efficiencies.

selected from the uninduced plates for reporter protein activity. Results show there is weak dose-dependent repression of the reporter protein upon IPTG induction (see Figure 3.26 and Table 3.8). In addition, analysis of the same cultures revealed that the new pGEX recombinants are expressing our proteins (see Figure 3.27), although it is not clear whether the proteins are soluble.

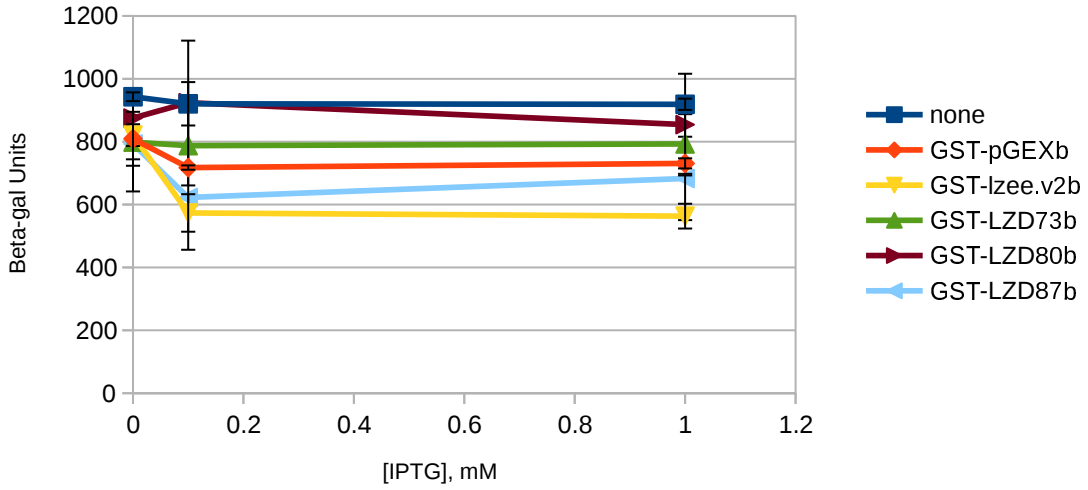


Figure 3.26: β-galactosidase levels from FW102/IC-450 reporter strains transformed with lacUV5-pGEXb (empty vector, lzeeV2, LZD73, LZD80, LZD87).

3.4.10 Reporter assays on pBAD recombinants

Lastly, recombinant pBAD plasmids were used as expression plasmids to test for DNA looping-dependent repression of β-galactosidase. Protein expression from these plasmids is regulated by the arabinose promoter, which should not interfere with the *lacZYA* expression. Furthermore, we have not seen any sequence homologies between the pBAD plasmids and F' episome.

Table 3.8: Reporter protein behavior in FW102/IC-450 looping strains transformed with pGEXb plasmids. There is no dose-dependent repression of the reporter protein upon expression of the gene inserts in pGEXb.

Expression Plasmid	β-gal Units (0 mM IPTG)	β-gal Units (1 mM IPTG)	Repression Ratio, RR
—	943 ± 14	919 ± 18	1.03 ± 0.03
pGEX6P1b	809 ± 86	731 ± 16	1.1 ± 0.1
lzeeV2b	821 ± 35	563 ± 39	1.5 ± 0.1
LZD73b	799 ± 157	793 ± 95	1.0 ± 0.2
LZD80b	875 ± 95	854 ± 162	1.0 ± 0.2
LZD87b	793 ± 49	683 ± 133	1.2 ± 0.2

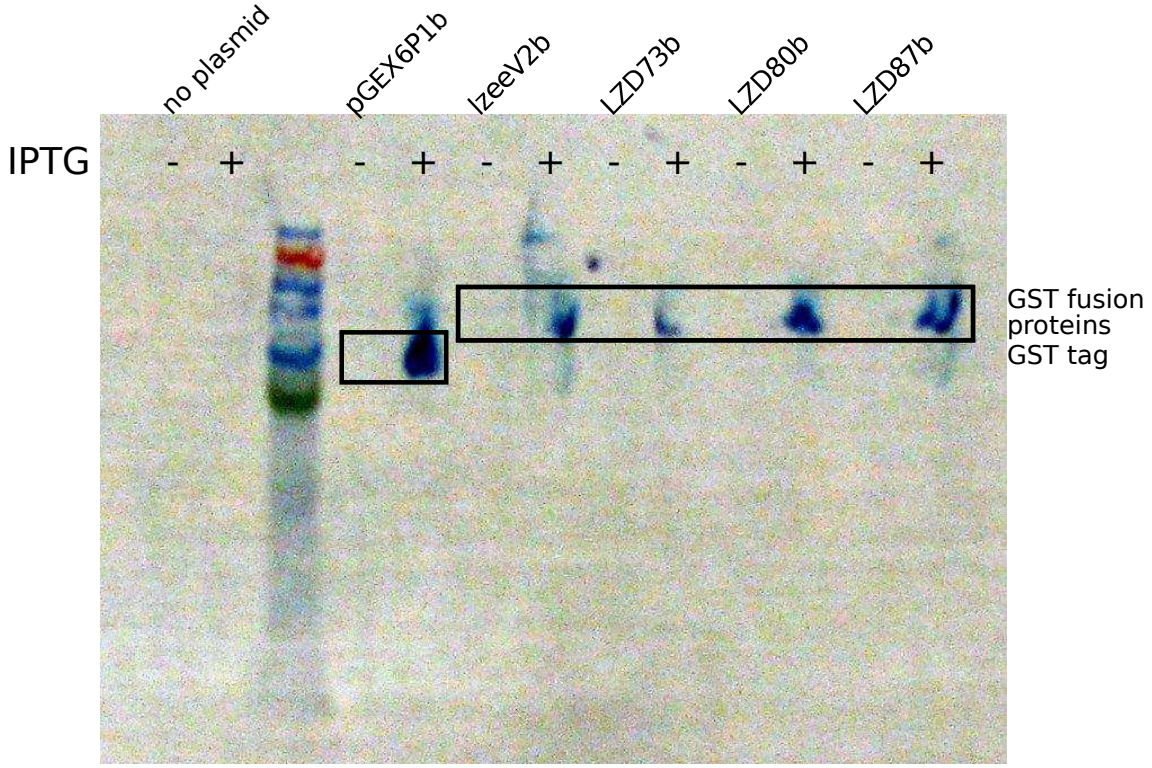


Figure 3.27: Western blot of crude lysates from lacUV5-pGEXb transformants (empty vector, lzeeV2, LZD73, LZD80, LZD87) in FW102/IC-450 reporter strains, 2 hours post-induction with 1 mM IPTG.

FW102/IC-450 transformed with the recombinant pBAD plasmids generated white to light pink colonies on TZ-glucose and TZ-lactose-arabinose plates. Trans-

formants were selected from both plates and analyzed for reporter protein activity. Results show that there is no significant repression of the β-galactosidase from any of the pBAD recombinants tested (see Figure 3.28 and Table 3.9).

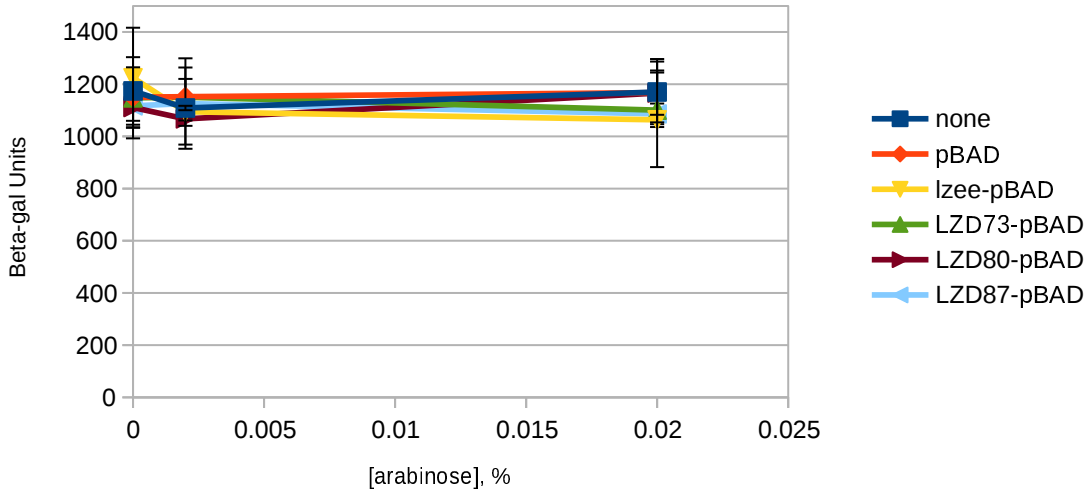


Figure 3.28: β-galactosidase levels from FW102/IC-450 reporter strains transformed with pBAD (empty vector, lzeeV2, LZD73, LZD80, LZD87).

A closer inspection of the genotypes of all our reporter strains showed a mutation in the *araD* gene in the F' episome, which prevents metabolism of arabinose by the host cell, but no mutation or deletion in *araEFGH* genes, which enables arabinose transport into the host cell (Whipple, 1998; Becker et al., 2005). Furthermore, SDS-PAGE analysis of whole cell extracts obtained from cell cultures used in the Miller assays did not show the presence of the putative proteins upon induction by arabinose (results not shown). This might indicate the absence, or if present, an inefficient one, of the arabinose transporter in the host cells.

Table 3.9: Reporter protein behavior in FW102/IC-450 looping strains transformed with pBAD plasmids. There is no dose-dependent repression of the reporter protein upon expression of the gene inserts in pBAD.

Expression Plasmid	β-gal Units (0 % arabinose)	β-gal Units (0.02 % arabinose)	Repression Ratio, RR
—	1174 ± 129	1170 ± 116	1.0 ± 0.1
pBAD	1150 ± 115	1168 ± 85	1.0 ± 0.1
lzeeV2	1225 ± 192	1064 ± 181	1.2 ± 0.3
LZD73	1148 ± 155	1101 ± 202	1.0 ± 0.2
LZD80	1111 ± 51	1166 ± 130	1.0 ± 0.1
LZD87	117 ± 124	1086 ± 40	1.0 ± 0.1

3.5 Conclusions

The genetic assay originally developed by Whipple was designed to study DNA-binding proteins from any source in bacterial cells (Whipple, 1998). Several applications of this assay to probe DNA flexibility by the use of a DNA-looping dependent repression has shown that non-specific DNA bending proteins, and the flexibility of the DNA looping protein itself contribute to the stability of very small DNA loops *in vivo* (Becker et al., 2005, 2007). In this chapter, we adapted the same assay to determine the contribution of inherent DNA flexibility in *in vivo* loop formation by using small and relatively rigid artificial DNA looping proteins that have been previously shown to loop DNA *in vitro* (Gowetski et al., 2013). In course of optimizing our genetic assays, we have observed either cell death, or permanent repression in reporter strains that contain relatively short distances between GCN4 sites upon expression from the pGEX recombinants. We proposed possible recombination events that have led to the loss of the F' episome in these strains, and have

seen some evidence to support this explanation.

On the other hand, we have observed relatively weak enhancement of repression on reporter strains with longer separation between GCN4 sites in the presence of our proteins. It is not clear at this point whether this enhancement is due to DNA looping. One future direction for this project is to do measurements on a much shorter distance range, that spans one helical turn of the DNA, to illustrate reporter protein repression that is dependent on DNA looping. By changing the relative orientations of the binding sites, we might expect to see periodic fluctuations in the calculated repression ratios.

Given the unexpected effects observed in shorter distances between GCN4 operators, it is possible that LZD proteins may be stabilizing much shorter DNA loops *in vivo* after all. However, its leaky expression in lacUV5-pGEX may trigger the recombination events in the host cells as a response to deal with the toxic protein. It would be interesting to note if there is a missing link between the expression of the LZD proteins and host cell toxicity that is a consequence of DNA looping.

Chapter 4: Catching DNA-protein loops using Plasmid Conformation Capture

4.1 Overview

Generating specialized *E. coli* cells that house the modified operator-promoter construct in single-copy for the reporter gene assay requires considerable time and effort. Moreover, it was a challenge selecting for transformants that survived overnight growth in liquid culture, which is required to measure β-galactosidase levels, as these looping proteins have been previously shown to be toxic to its cell host when expressed in high amounts. Thus, another assay was developed to directly analyze various operator distances for DNA loop formation with the LZD proteins. The chromosome conformation capture (3C) method, developed by Dekker and colleagues, was adapted for this system (Dekker, Rippe, Dekker, & Kleckner, 2002). While the reporter gene assay looks at changes in reporter protein levels in the presence of a DNA looping protein, a consequence of DNA looping, 3C offers a more direct method of analyzing protein-mediated DNA loops in *in vivo* and *in vitro* conditions. For this system, this method will be renamed plasmid conformation capture (PCC) since the binding sites for the LZD proteins are cloned into a low-copy plasmid

instead of the bacterial chromosome.

4.2 Rationale

In this assay, whole cells or solutions containing the artificial looping protein and the DNA binding sites were fixed using formaldehyde, which enables formation of cross-links between the protein and the DNA. The DNA in the covalently linked protein-DNA complex is subsequently digested by a restriction enzyme, followed by a ligation step at very low DNA concentrations that favors intramolecular ligation of DNA fragments cross-linked to the same protein. The cross-links were then removed and specific ligation products were quantified using polymerase chain reaction. The amount of each ligation product is indicative of the nature and frequency of interaction between DNA fragments (see Figure 4.1).

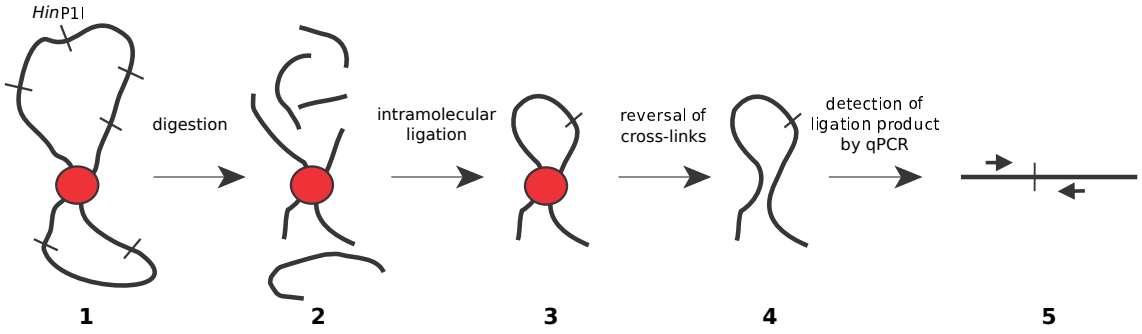


Figure 4.1: Plasmid Chromosome Capture. (1) In this method, host cells expressing the LZD proteins (red circle) and the plasmid DNA containing binding sites for the protein is fixed with formaldehyde, making protein-protein and protein-DNA cross-links. (2) The cross-linked complex is digested with a restriction enzyme (*Hin*P1I) and (3) ligated at very low concentrations. (4) The cross-links are subsequently removed and (5) the new ligation junction is detected by PCR.

As outlined in Figure 4.1, a low-MW PCR product should only form upon ligation of two DNA fragments. The amount of a specific ligation product depends

on the local concentration and orientation of the DNA termini. If there is a direct interaction between two DNA loci, the efficiency of ligation increases. A specific interaction occurs if the following criteria are met: (1) bimolecular ligation product yields from the same DNA fragments are low in the absence of the looping protein, and (2) ligation product yields increase significantly upon crosslinking of the looping protein to both DNA fragments. Thus, a protein-DNA complex can be detected if there is a direct interaction between the looping protein and two DNA fragments containing binding sites for the looping protein.

To generate DNA templates for PCC, the multiple cloning site in the transfer plasmids containing the operator-promoter construct in the reporter assay is extended by inserting a 1100-bp DNA fragment derived from lambda DNA between the *Eco*RI and *Sac*I recognition sites. This allows the subsequent cloning of the INV2 (5'-GTCATATGAC-3') operator, a strong binding site to the C-terminal DNA binding domain of the LZD proteins, at upstream positions ranging from 300 to 1150 base pairs from the proximal GCN4 site by site-directed mutagenesis (see Chapter 3.3.1.4). Next, a restriction enzyme is chosen for the digestion step that enables high resolution and near coverage of the region of interest. In this system, very short-range interactions are being mapped out compared to 3C methods. Thus, a restriction enzyme that would generate at least 100-bp digestion fragments is picked. *Hin*P1I and *Cvi*QI are selected as possible candidates for this system. PCR primers are then designed on the region of interest, based on the location of the restriction enzyme cut sites. In addition, internal control primers located outside the region of interest are designed for detecting intermolecular ligation, which is independent on

LZD binding to its DNA site/s in the plasmid, and normalization primers that determine template yields. A summary of DNA templates and PCR primers designed, based on the location of *Hin*P1I sites, for this assay is shown in Figure 4.2.

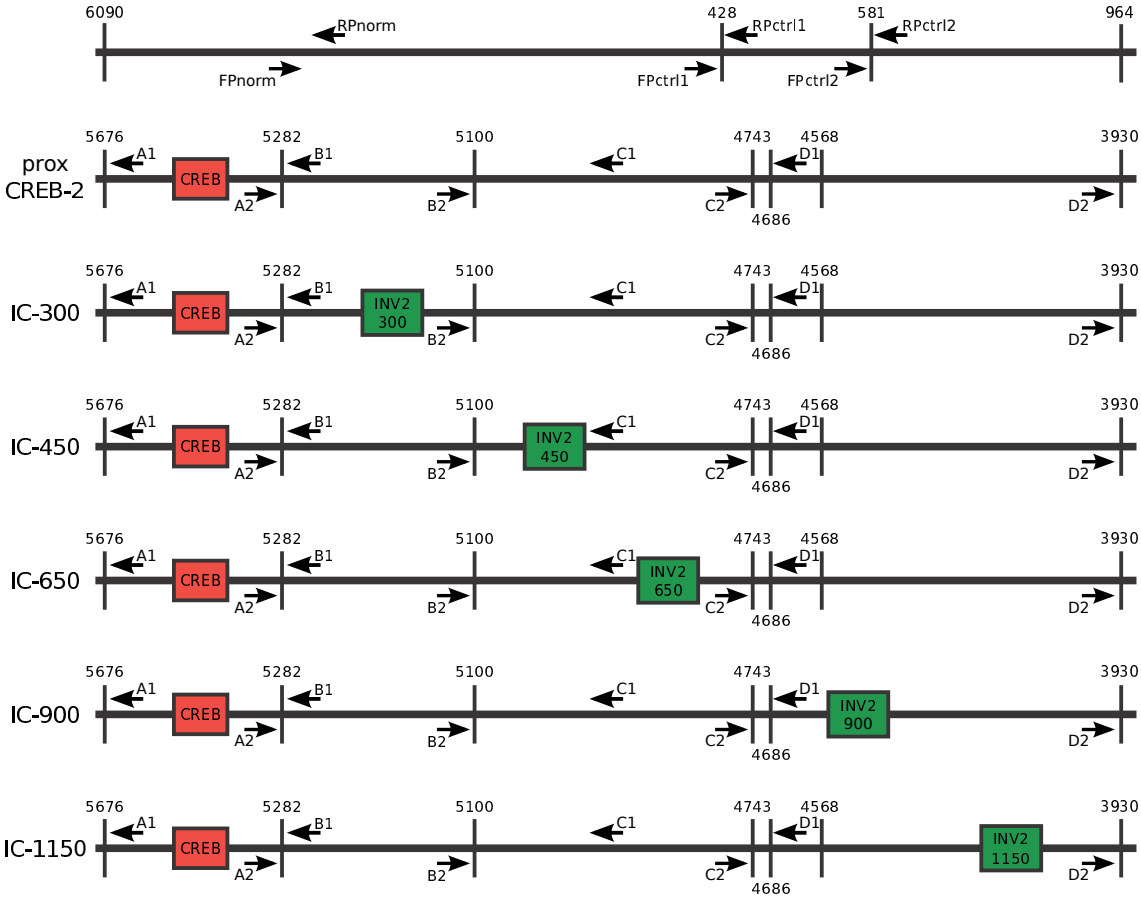


Figure 4.2: DNA templates and PCR primers used in PCC. Position of GCN4 binding sites (CREB in red boxes, INV2 in green boxes) are not drawn to scale. Vertical lines indicate *Hin*P1I cut sites. PCR primers FPnorm and RPnorm determine DNA template yield, while any combination of PCR primers FPctrl1, FPctrl2, RPctrl1, and RPctrl2 detects intermolecular ligation. Sequences of all the primers shown in this figure are listed in Appendix 2.5

.

Several factors must be taken into consideration in order to make a correct analysis of results. First, template amounts have to be determined among all samples such that the amount of a specific PCR product increases linearly with increasing

amounts of template. Second, a control template should be generated such that it contains equal amounts of all possible ligation products to account for differences in primer efficiencies. Third, the level of background random interactions has to be determined by testing several distances farther along the region of interest from the DNA fragment containing the primary operator. A looping interaction would be identified if there is a local peak in the interaction frequency between DNA fragments containing the binding sites. Fourth, another set of control PCR primers have to be designed in a different region of the plasmid to account for differences in cell preparations and experiment conditions. These internal controls would determine whether a specific interaction is correlated with biological activity such as the expression of a DNA looping protein. Finally, in order to account for differences in DNA yields, and interaction frequencies from various sample preparations, the amount of PCR product from a specific ligation product is normalized to both the PCR product that does not amplify across a ligation junction, and to a PCR product from a bimolecular ligation event that is unaffected by the looping protein being studied, respectively (Dekker, 2006).

In this study, we define the interaction frequency of a DNA ± protein sample as the ratio of concentrations of a specific ligation product, e.g. using primers A1 and B2, to that of the internal control primers. Furthermore, the crosslinking efficiency is defined as the ratio of interaction frequency between formaldehyde-treated to uncrosslinked samples. To account for differences in sample preparation from different experiment conditions, the interaction frequencies are normalized against the DNA yield, which is quantified by PCR using FPnorm and RPnorm primers.

4.3 Materials and Methods

Reagents and enzymes for molecular biology work were purchased from NEB, while DNA oligonucleotides for PCR were purchased either from IDT or Sigma. Quantitative PCR (qPCR) was done on a LightCycler® 480 Real-time PCR Instrument (Roche).

Promoter-operator constructs prepared as described in Chapter 3.3.1.4 were used as the plasmid templates for PCC. The LZD expression plasmids used for *in vivo* PCC are the recombinant lacUV5-pGEX plasmids.

4.3.1 *In vivo* PCC

We have adapted and modified the procedure developed by Dekker and colleagues (Miele, Gheldof, Tabuchi, Dostie, & Dekker, 2006) to work on *E. coli* samples. First, TOP10 *E. coli* cells were transformed with the plasmid DNA containing the GCN4 operators and the expression plasmid containing the LZD protein. Correct transformants were identified using colony PCR that detects both the intact operators (seqFP-OdOp + Op-rp2a) and the LZD gene sequence (pGEXfp3 + pGEXrp), and then grown in LB media containing amp, kan, cam, and strep. The seed cultures were then split up and diluted 50-fold into fresh growth media ± 1 mM IPTG and grown to mid-log phase. An aliquot of liquid culture from each setup was then treated with 1% formaldehyde at room temperature for five minutes, and quenched with 0.5 M glycine at 4 °C for 10 minutes. The cells were then harvested after the cross-linking step. Cells from the remaining aliquots were harvested and

used as the uncross-linked control samples. The cells were lysed at 37 °C for 30 minutes without agitation in Lysis Buffer (10 mM Tris-HCl pH 8.0, 20% sucrose, 50 mM NaCl, 10 mM EDTA, 10 mg/mL lysozyme), and the lysates could then be stored at -80 °C upon addition of 1 mM PMSF, or 1X protease inhibitor cocktail.

An aliquot of the lysate was washed three times in the restriction enzyme buffer and treated with 0.1% SDS for 10 minutes at 65 °C to remove any uncross-linked proteins. Excess SDS was then sequestered with 1% Triton X-100 before incubating the DNA in 1.2 U/μL restriction enzyme for at least 8 hours. The restriction enzyme was then inactivated by incubating the digested samples with 1.5% SDS for 20 minutes at 65 °C. The digestion products were diluted 15-fold with 1X T4 DNA Ligase Buffer supplemented with 1% Triton X-100, 1 mM ATP and 100 μg/mL BSA, and then ligated for 2 hours at 16 °C with 1 U/μL T4 DNA Ligase. Cross-links and all proteins were then removed by incubating the ligation products with 50 μg Proteinase K at 65 °C for at least 12 hours. Finally, DNA was isolated and worked up by phenol/chloroform extraction and ethanol precipitation, and the DNA pellets resuspended in TE pH 8 buffer. Control DNA samples – (1) intact plasmid, (2) digested DNA, and (3) randomly ligated DNA – were isolated from each step of the method. Specific ligation products were then analyzed by PCR and the resulting PCR products were visualized using agarose gel electrophoresis.

4.3.2 *In vitro* PCC

Plasmid DNA containing GCN4 binding sites (500 pM) were incubated with LZD protein (25 nM) at room temperature for at least 30 minutes. Afterwards, an aliquot of this mixture was treated with 0.05% formaldehyde at 4 °C for 20 minutes, and then quenched with 0.5 M glycine at 4 °C for 10 minutes. The protein-DNA complexes were diluted 10-fold with the restriction enzyme buffer before incubating in 1.2 U/μL restriction enzyme for at least 8 hours. The restriction enzyme was then inactivated with 1.5% SDS at 65 °C for 20 minutes, then diluted the samples 15-fold with 1X T4 DNA Ligase Buffer supplemented with 1% Triton X-100, 1 mM ATP and 100 μg/mL BSA, and ligated with 1-1.2 U/μL T4 DNA Ligase for two hours at 16 °C. Cross-links and all other proteins were removed by incubating at 65 °C for at least 12 hours with 50 μg Proteinase K. The ligation products were then isolated and worked up by phenol/chloroform extraction and ethanol precipitation, and the DNA pellets resuspended in TE pH 8.0 buffer. Control samples from both uncross-linked and cross-linked samples were also isolated from each step of the experiment. Finally, specific ligation products were analyzed by PCR.

4.3.3 Quantitative PCR (qPCR)

The qPCR master mix (20 μL per reaction) that contained the following was prepared: 1X SYBR Green I Mix (proprietary mixture containing FastStart Taq DNA Polymerase, reaction buffer, dNTP mix with dUTP replacing dTTP, SYBR Green I dye and $MgCl_2$), 250 nM each of the qPCR primers, and 1-5 μL DNA

template.

The following program was set up on the thermal cycler:

- Initial Denaturation – 94 °C at 4.80 °C/s; hold for 2 minutes
- Amplification (35 cycles)
  - 94 °C at 4.80 °C/s; hold for 15 seconds
  - 60 °C at 2.50 °C/s; hold for 15 seconds
  - 68 °C at 4.80 °C/s; hold for 30 seconds, single acquisition mode
- Melting Curve
  - 95 °C at 4.80 °C/s; hold for 5 seconds
  - 65 °C at 2.50 °C/s; hold for 1 minute
  - 97 °C at 0.11 °C/s; continuous acquisition mode (five acquisitions per °C)
- Cooling – 40 °C at 2.50 °C/s; hold for 30 seconds

4.3.4 qPCR data analysis

First, a concentration curve was generated for each PCC plasmid template by plotting the cycle threshold values, C_t, which were extrapolated from the qPCR results with FPnorm+RPnorm primers, against the logarithm of concentrations from serial dilutions of undigested plasmid minipreps. The best fit line from this curve was then used to calculate the concentrations of the PCC templates used in all qPCRs.

The calculated concentrations of the PCC templates from qPCRs using A1+B2, A1+C2 and A1+D2 primers were normalized against concentrations derived from qPCRs using FPnorm+RPnorm or FPctrl1+FPctrl2 primers. Lastly, the crosslinking efficiency is calculated as follows: [normalized PCR product]$_{+\ \text{formaldehyde}}$/[normalized PCR product]$_{-\ \text{formaldehyde}}$.

4.4 Results and Discussion

4.4.1 *In vivo* PCC control experiments

Optimization of existing 3C protocols to our system was done *in vivo* using the plasmids proxCREB-2 and IC-300, the expression plasmids (lacUV5-pGEX6P1) for LZD80 and LZD87, and *Hin*P1I as the restriction enzyme. PCR was initially done on undiluted templates from all cell preparations previously mentioned, to check for the efficiencies of restriction enzyme digestion and religation under reaction conditions in the capture experiment. Figures 4.3 and 4.4 showed that the 217-bp PCR product from primers A1 and B2 only showed up in samples that were digested and successfully ligated. The results indicate that the cross-linking stringency and template concentration are low enough for a near-complete digestion of cross-linked plasmid DNA. Moreover, cross-linking seems to increase the amount of ligation product.

PCR was then done on serial dilutions of the proxCREB-2 templates (+ IPTG) using primers A1 and B2 (see Figure 4.5) to optimize template concentrations to the linear range. Results on ligation control DNA showed a significant decrease in

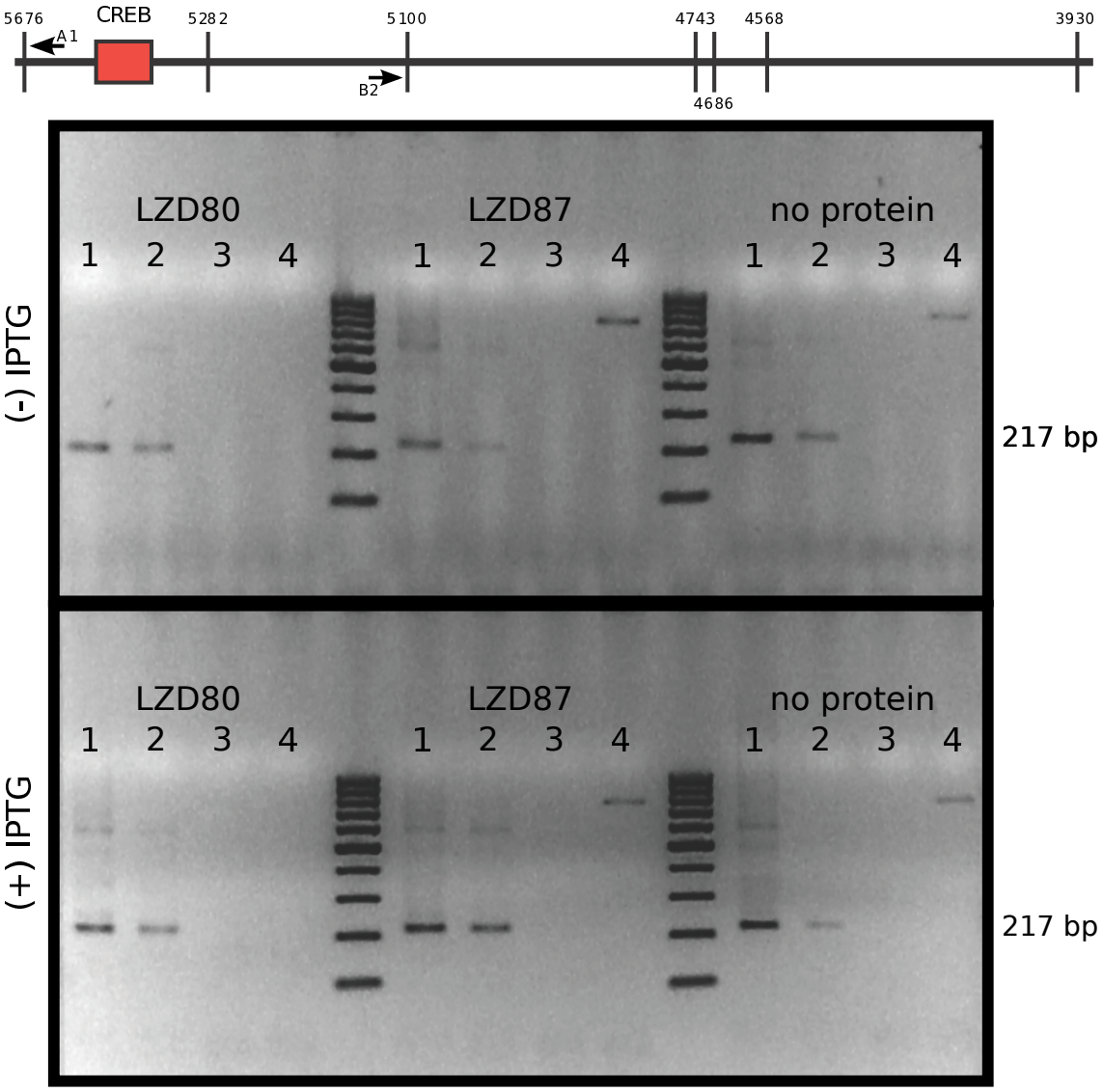


Figure 4.3: *In vivo* PCC control experiments for proxCREB-2 ± LZD80 or LZD87. Gel electrophoresis (1.5 % agarose in TBE) of PCR products obtained with primers A1 and B2. 1 – PCC template, 2 – ligation control (cross-links were removed prior to ligation step), 3 – digestion control, 4 – undigested template

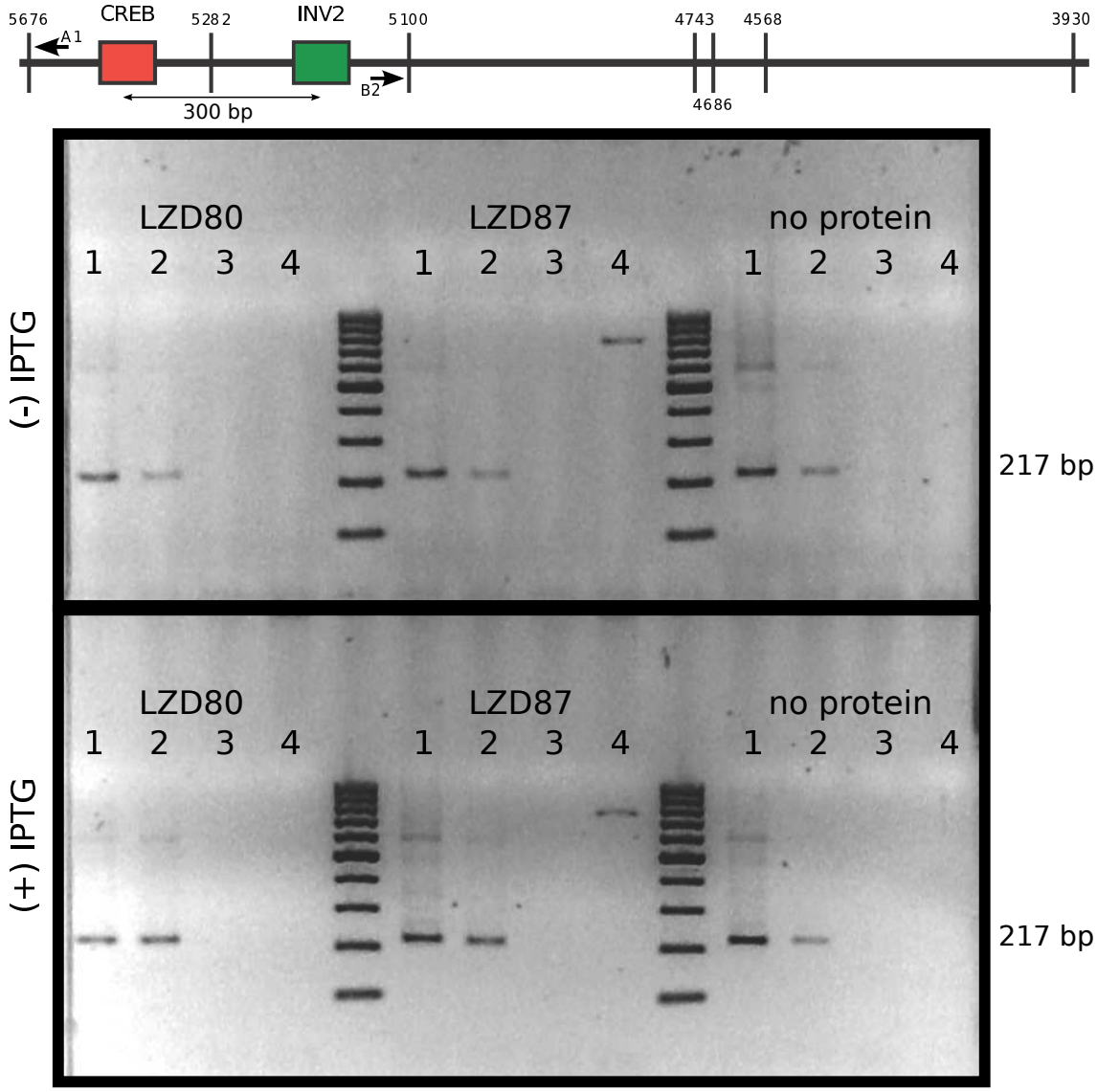


Figure 4.4: *In vivo* PCC control experiments for IC-300 ± LZD80 or LZD87. Gel electrophoresis (1.5 % agarose in TBE) of PCR products obtained with primers A1 and B2. 1 – PCC template, 2 – ligation control (cross-links were removed prior to ligation step), 3 – digestion control, 4 – undigested template

the amount of PCR product, compared to the more gradual decrease of the PCR product from the PCC template. The results show that intermolecular ligation products are not detectable in the linear range of template concentration for PCR.

Next, we proceeded to quantifying the major PCR products by measuring band intensities in a polyacrylamide gel stained with SYBR Gold (Life Technologies). However, the stained gel revealed the presence of comparable amounts of higher MW PCR products to the major PCR product in the PCC samples (see Figures 4.6 and 4.7). Apparently, the digestion efficiency was not nearly as complete as we have previously thought. We attribute this to residual formaldehyde in the samples that partially inactivated the restriction enzyme, which in turn, resulted to incomplete digestion of our templates.

Based on the results presented so far, the assay seems to be working although the biggest deterrent to continue with the quantification of specific ligation products is the trace amounts of formaldehyde in the cross-linked samples, which affects the amount of specific ligation products being quantified. We have done the experiment using much higher amounts of restriction enzyme and ligase to improve digestion and ligation efficiencies, but due to the sheer number of templates to be analyzed and experiment conditions to consider, this assay can be quite expensive in the long run. Thus, we need to review and modify the sample preparation method prior to enzyme treatment of these samples.

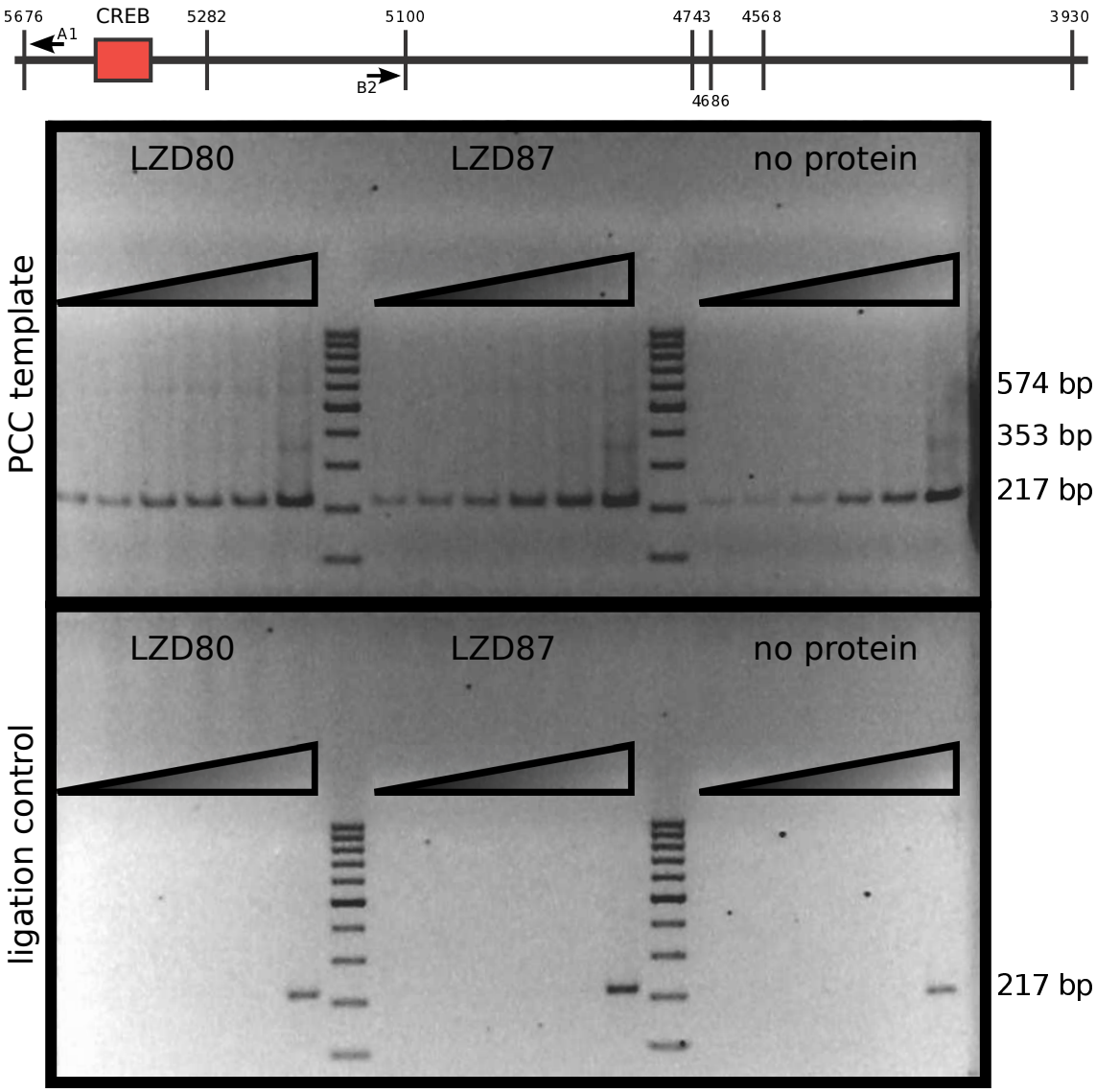


Figure 4.5: Template dilutions of proxCREB-2 ± LZD80 or LZD87. Gel electrophoresis of PCR products obtained with primers A1 and B2, using serial dilutions of the DNA template (from left to right: 1600-fold, 800-fold, 400-fold, 200-fold, 100-fold, undiluted template). Results indicate that intermolecular ligation is undetectable at 100-fold dilution of the PCC template

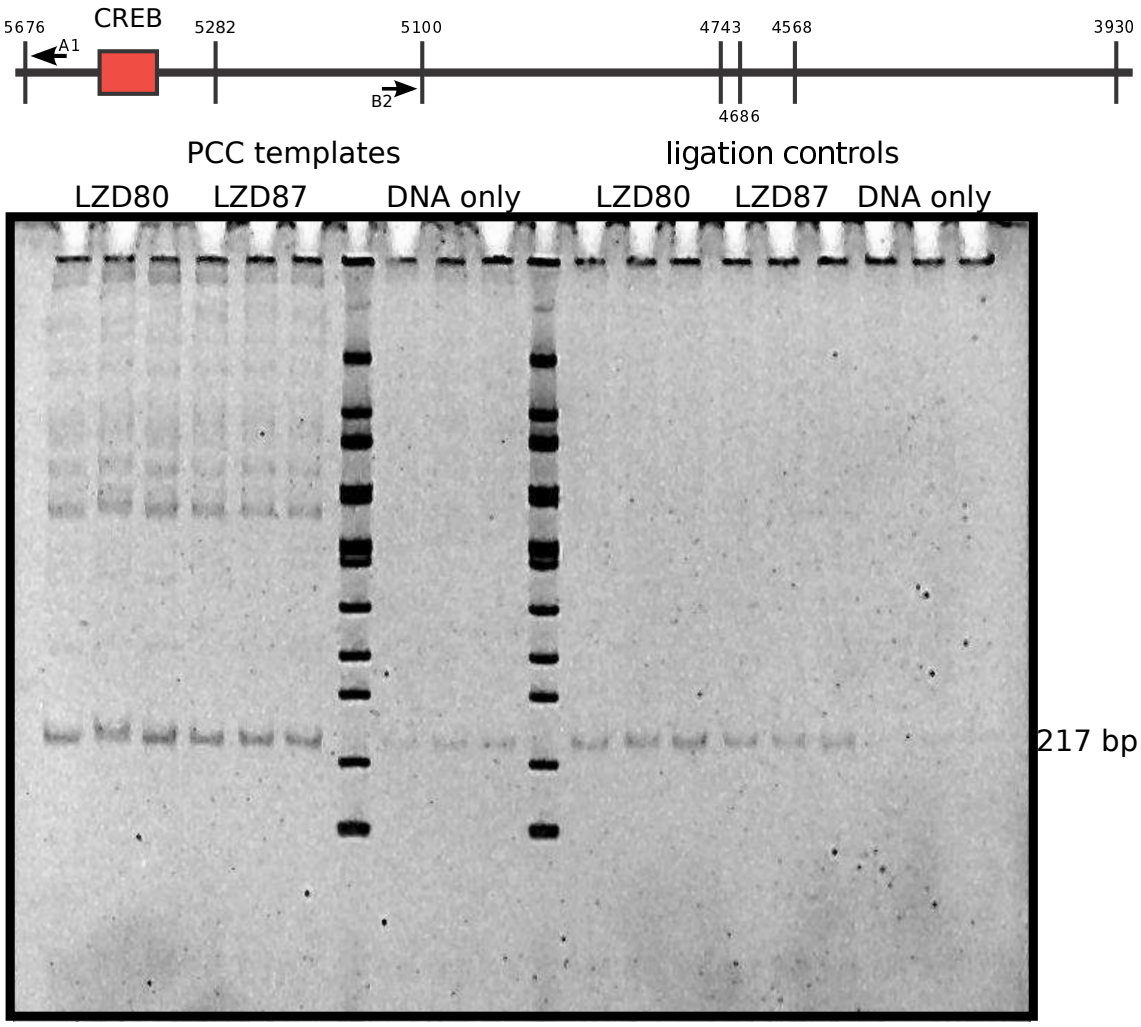


Figure 4.6: Quantification of *in vivo* PCC A1+B2 PCR products in proxCREB-2 ± LZD80 or LZD87 + IPTG. Native PAGE of PCR products obtained with primers A1 and B2. PCR was done in triplicate per template.

4.4.2 *In vitro* PCC control experiments

In order to simplify the analysis of our *in vivo* PCC results, we performed the assay *in vitro* using plasmid DNA and purified protein. First, we started out with a less stringent crosslinking condition (i.e. lower [CH_2O] and temperature) to give us enough crosslinked product but not adversely affect restriction enzyme digestion and religation activities. PCR was done with internal control primers

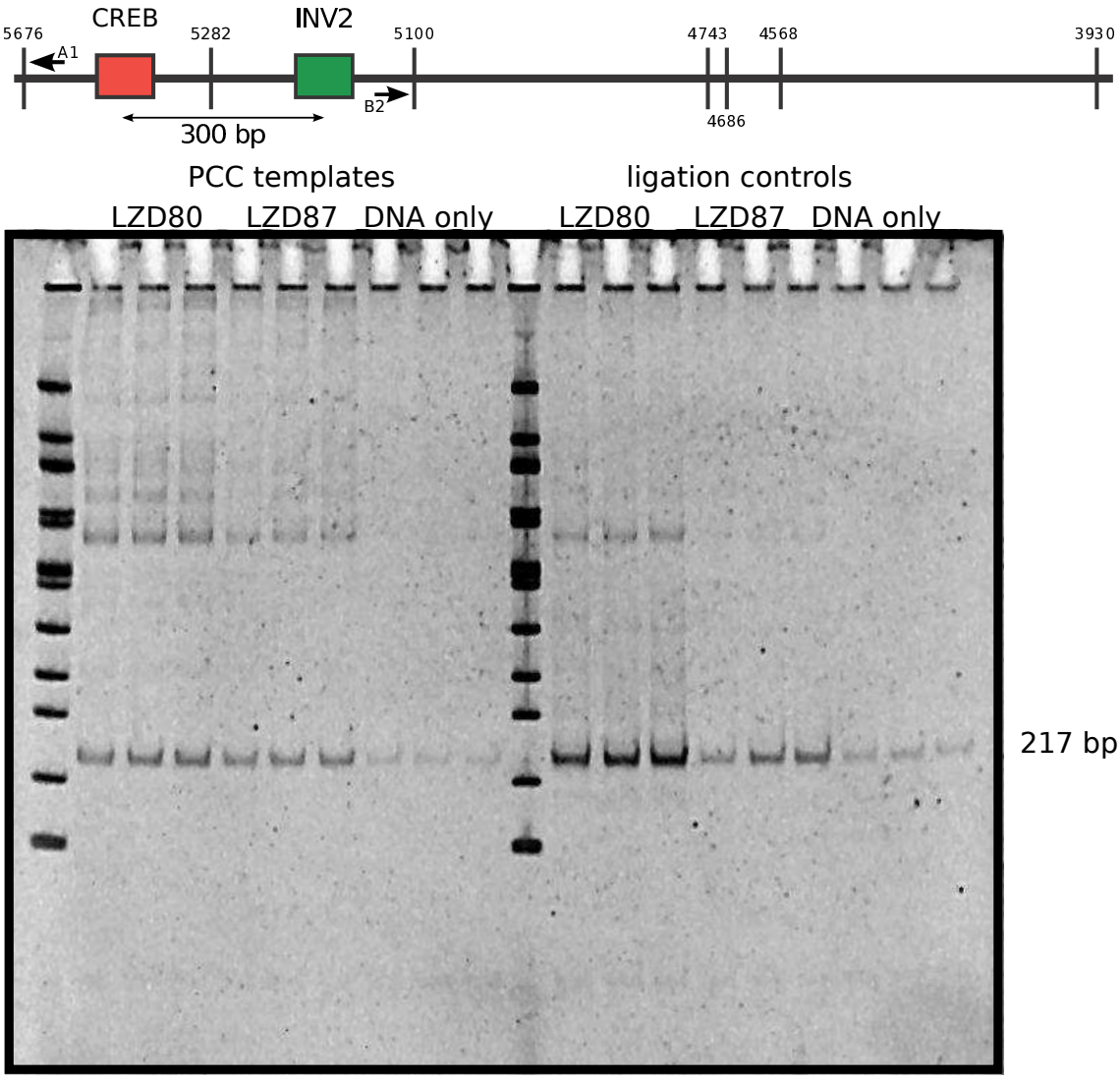


Figure 4.7: Quantification of *in vivo* PCC A1+B2 PCR products in IC-300 ± LZD80 or LZD87 + IPTG. Native PAGE of PCR products obtained with primers A1 and B2. PCR was done in triplicate per template.

FPctrl1 and RPctrl2. Results are shown in Figure 4.8. The 156-bp product only shows up in the PCC samples and the ligation controls, indicating that the *Hin*P1I sites at 428 and 581 have been cut and religated. The amount of this product reflects the extent of intermolecular ligation. On the other hand, the amount of the 309-bp product reflects the amount of the PCR template in the intact plasmids, the

extent of digestion in the digested samples, and the amount recovered from uncut and incompletely digested plasmids upon ligation. For uncross-linked templates, the relative amounts of both PCR products have not been significantly enhanced with the addition of the LZD protein. This is to be expected since there are no GCN4 binding sites in that region. Formaldehyde treatment prior to digestion and ligation resulted to a slight increase in the amount of the 309-bp PCR product in the PCC templates, which is most likely coming from the amount of cross-linked samples that could not be digested and were recovered upon the removal of the cross-links.

Next, we checked to see if there is a direct interaction between the GCN4 sites in the IC-450 template upon LZD crosslinking. PCR results using primers A1 and C2 showed the emergence of a 222-bp PCR product upon religation of *Hin*P1I fragments containing CREB or INV2 operators (see Figure 4.9). This PCR product is more pronounced when the LZD protein is cross-linked to the DNA template prior to digestion and ligation. On the other hand, the high-MW PCR products present in PCC templates containing the LZD proteins could be due to incomplete digestion of the template. This can be minimized by either doing the ligation at an even lower concentration of DNA, or by doing the PCR using a lower amount of DNA template.

From the results presented, it looks like the template concentrations have been optimized to amounts where we can detect specific ligation products upon crosslinking. Thus, we can proceed to quantifying the various specific ligation products in our DNA templates.

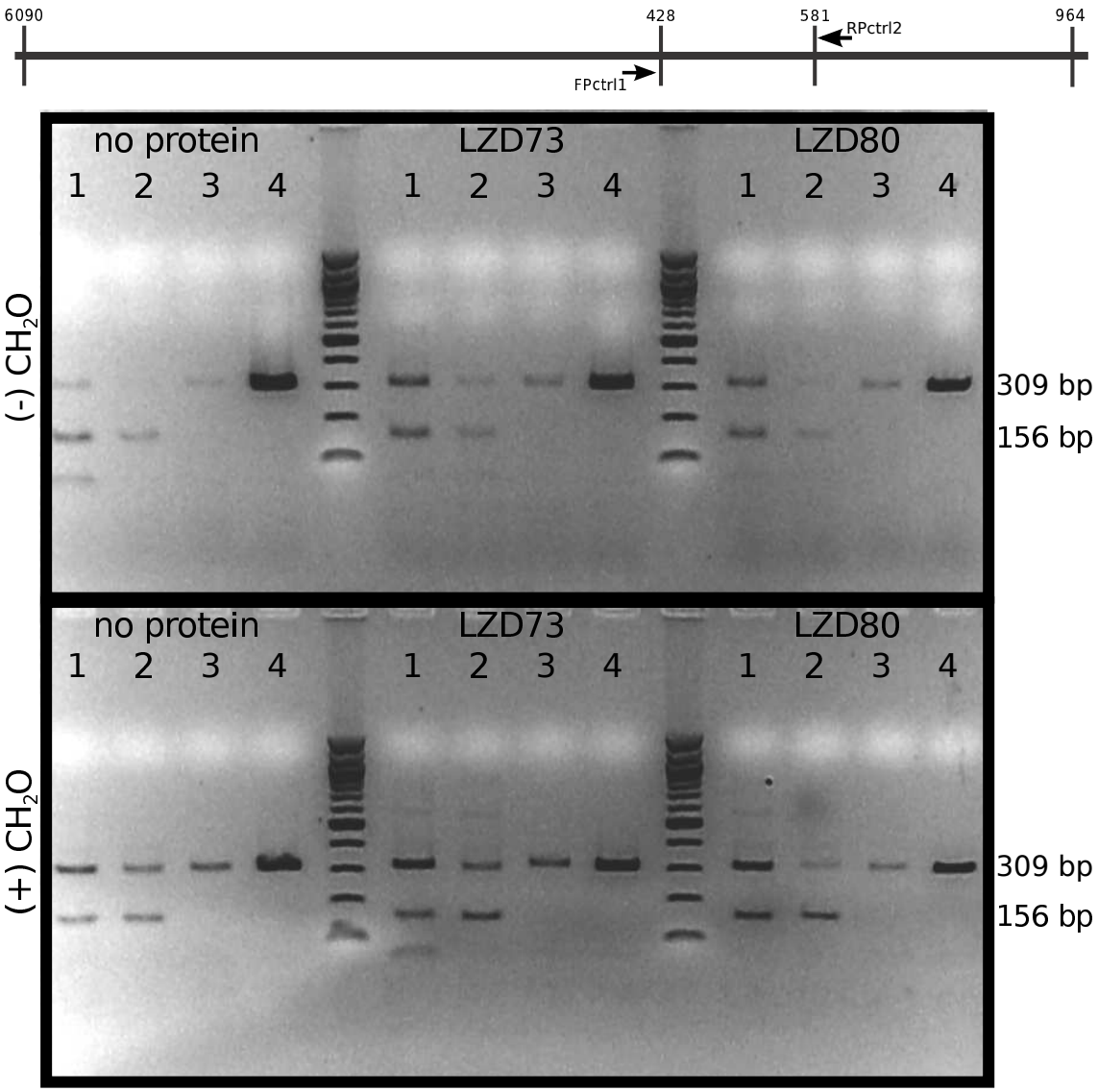


Figure 4.8: *In vitro* PCC control experiments for IC-450 ± LZD73 or LZD80. Gel electrophoresis (1.5 % agarose in TBE) of PCR products obtained with internal control primers FPctrl1 and RPctrl2. Results highlight the appearance of the 156-bp PCR on templates that have been successfully digested and ligated. 1 – PCC template, 2 – ligation control (cross-links were removed prior to ligation step), 3 – digestion control, 4 – undigested template

4.4.3 Quantification of *in vitro* PCC ligation products

In a parallel experiment, we attempted to concentrate His-tagged LZD73-DNA complexes from the *in vitro* binding and crosslinking reactions prior to the *Hin*P1I

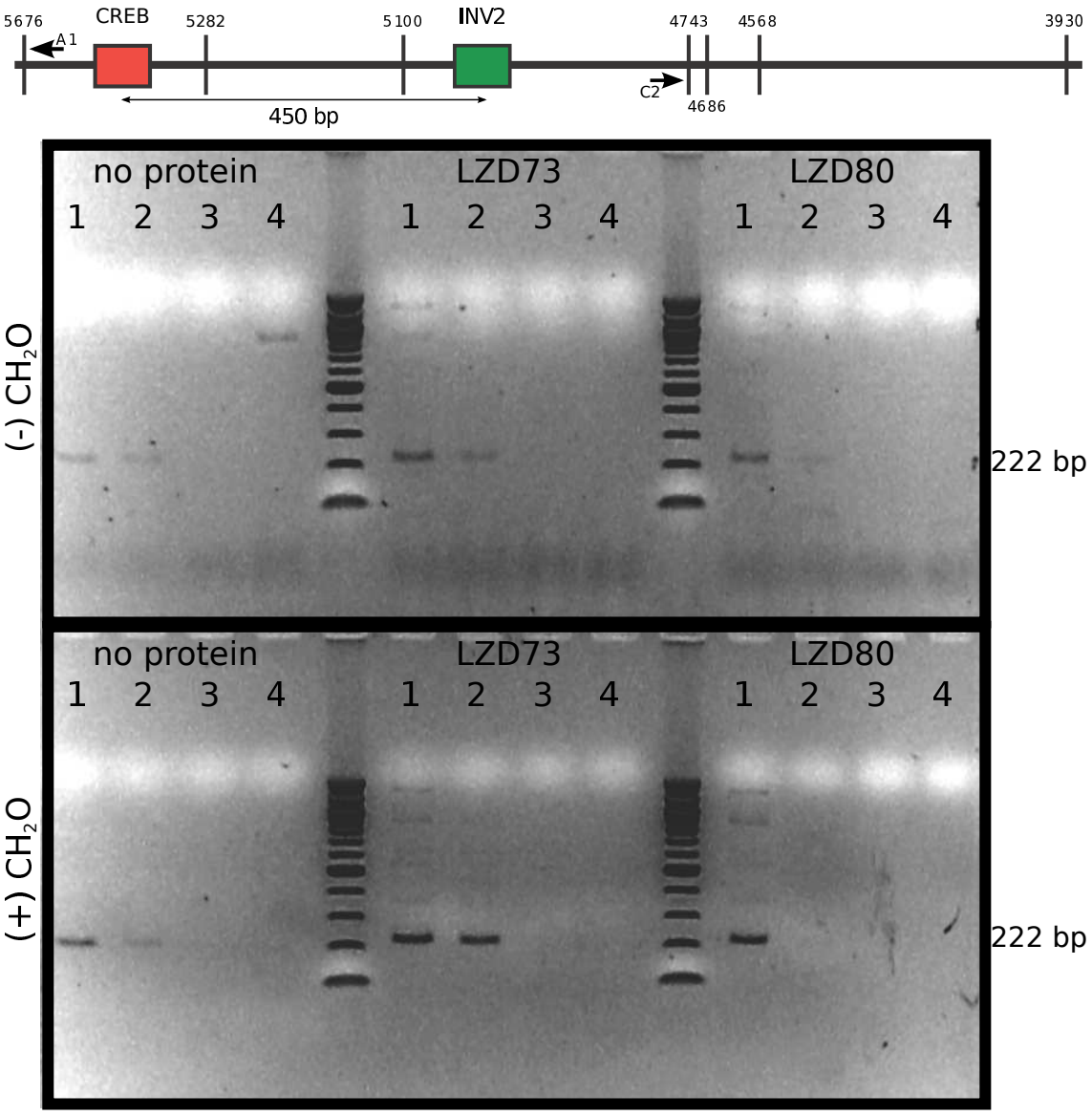


Figure 4.9: *In vitro* PCC control experiments for IC-450 ± LZD73 or LZD80. Gel electrophoresis (1.5 % agarose in TBE) of PCR products obtained with primers A1 and C2. 1 – PCC template, 2 – ligation control (cross-links were removed prior to ligation step), 3 – digestion control, 4 – undigested template

digestion step using batch purification with Ni Sepharose FF beads. This additional step would have the advantage of isolating the cross-linked protein-DNA complex prior to enzymatic treatments, and would be expected to give a better signal for enhanced ligation on fragments containing the GCN4 sites without having to dilute the samples to minimize extent of intermolecular ligation. Specific PCR products from different plasmid DNA samples using various primer pairs were then detected using ethidium bromide on a 1.5% agarose gel (see Figure 4.10). Ligation products from fractions eluted with 500 mM imidazole (E1) cannot be detected with ethidium bromide, whereas those from the original mixture (crude) and fractions collected with 25 mM imidazole (S1) seem to have comparable amounts.

Quantitative PCR using FPnorm and RPnorm primers was then used to determine differences in the template amounts among the different samples. To do this, calibration curves were first constructed for each template using known concentrations of the intact plasmid templates. (see Figure 4.11).

We then used the concentration curves to determine the template yields in all our PCC templates (see Table 4.1).

The results showed that there is a 1.5- to 2.5-fold difference in the DNA yields between the crude and S1 samples, whereas that difference increases 100- to 400-fold between crude and E1 samples. This could indicate that (1) formaldehyde may not have cross-linked LZD to the plasmid successfully, (2) the crosslinking step had somehow affected the binding affinity of the Histidine tag in the LZD protein to the affinity beads, or (3) the DNA-protein complexes were not completely eluted out of the affinity beads. However, there is a correlation, albeit miniscule, in the

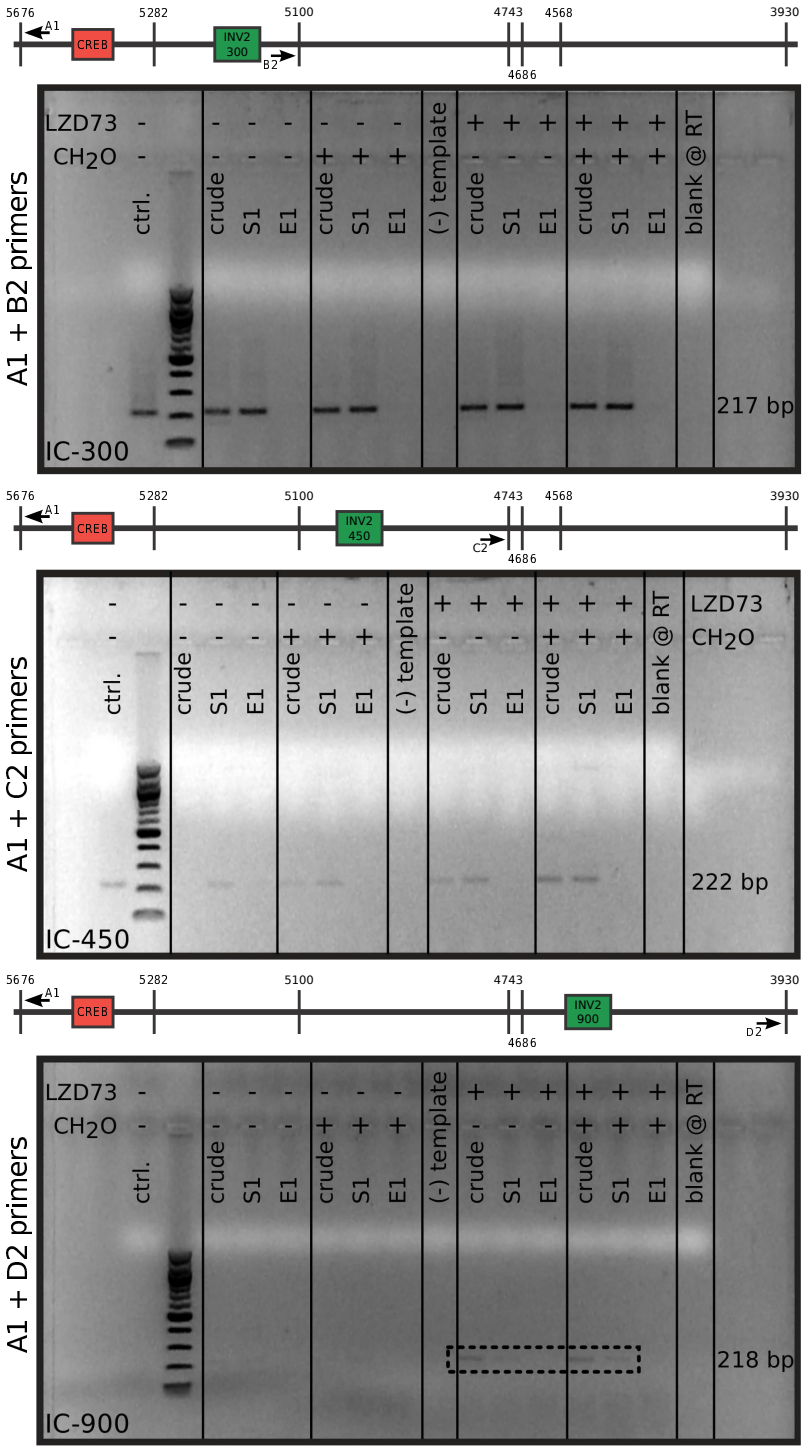


Figure 4.10: *In vitro* PCC for detecting 300-, 450- and 900-bp interaction between GCN4 sites using A1+B2, A1+C2, and A1+D2 primers, respectively. In a parallel experiment, enrichments of the crosslinked and non-crosslinked LZD-DNA complexes were done using batch purification on Ni Sepharose FF beads, which bind to 6x His tag fused to the LZD proteins. The dotted box highlights the presence of a very small amount of PCR product using A1+D2 primers. D2 primer has been redesigned.

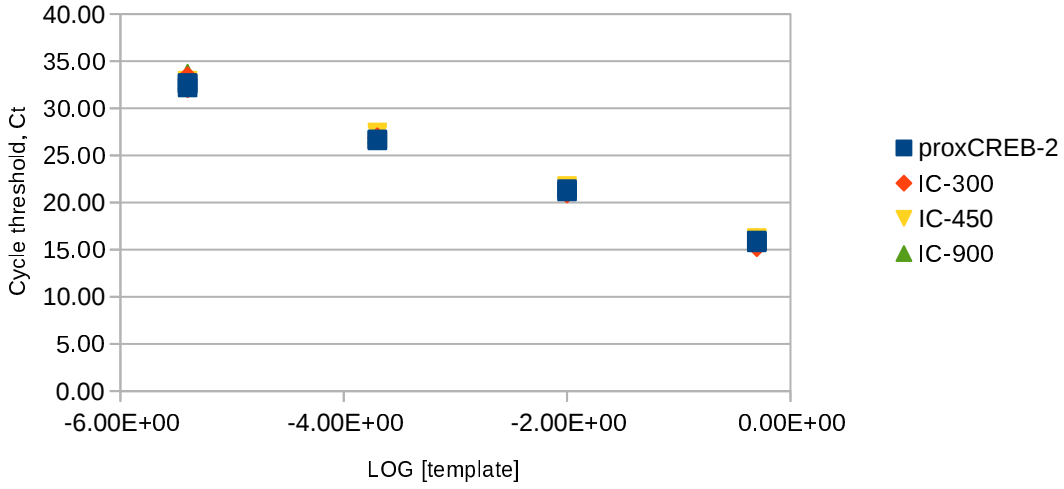


Figure 4.11: Concentration curves for *in vitro* PCC. Cycle threshold values, Ct, were extrapolated from three qPCRs per template concentration used.

DNA yields among E1 samples in all of the plasmid DNA used; there is a 2-to 5-fold increase in the template amounts when LZD is present, but an almost imperceptible change in DNA yield in the presence of formaldehyde. Despite the very small yields, we proceeded in quantifying ligation products that detect a 300-, 450-, and 900-bp interaction between GCN4 binding sites, using the crude samples as our DNA templates.

Table 4.2 summarizes the cycle threshold values extrapolated from an average of three qPCR runs for each sample prep. The concentrations of the PCR products were then calculated using the concentration curves in Figure 4.11. In order to account for differences in sample preparation and experiment conditions, the concentrations were normalized to (1) the amount of intermolecular ligation product (FPctrl1+FPctrl2 PCR product), or (2) the template yield (FPnorm+RPnorm

Table 4.1: Comparison of DNA yields in *in vitro* PCC templates, from qPCR data with FPnorm + RPnorm primers. Crude – PCC template, S1 – supernatant collected after incubating the LZD73-DNA mixture in Ni Sepharose FF beads, E1 – supernatant eluted with 500 mM imidazole

DNA	LZD73	CH_2O	C_t			[Template yield], pM		
			Crude	S1	E1	Crude	S1	E1
CREB	-	-	16.65	17.65	24.39	2.70	1.33	1.13E-02
CREB	-	+	16.75	17.85	24.76	2.51	1.16	8.73E-03
CREB	+	-	16.61	18.49	24.40	2.78	1.47	1.14E-02
CREB	+	+	16.58	17.89	23.72	2.84	1.12	1.82E-02
IC-300	-	-	16.70	17.63	25.71	1.99	1.06	4.65E-03
IC-300	-	+	17.16	17.83	25.38	1.45	0.93	5.84E-03
IC-300	+	-	16.86	17.51	24.07	1.78	1.15	1.41E-02
IC-300	+	+	16.85	17.54	23.60	1.80	1.13	1.92E-02
IC-450	-	-	17.06	17.65	25.49	2.59	1.70	7.06E-03
IC-450	-	+	16.63	17.78	24.84	3.50	1.56	1.11E-02
IC-450	+	-	16.72	17.44	24.05	3.27	1.99	1.93E-02
IC-450	+	+	16.90	17.43	23.79	2.90	2.00	2.31E-02
IC-900	-	-	17.54	18.39	25.90	1.82	1.01	5.37E-03
IC-900	-	+	17.16	17.94	25.73	2.37	1.38	6.05E-03
IC-900	+	-	17.01	18.01	23.61	2.63	1.31	2.64E-02
IC-900	+	+	17.13	17.90	23.98	2.42	1.41	2.04E-02

PCR product). Finally, the crosslinking efficiency for each normalized specific PCR product is calculated from the ratio between crosslinked and uncrosslinked samples.

The resulting crosslinking efficiencies from the different plasmid samples were overlaid, and plotted against the DNA length (see Figure 4.12). Since the effectiveness of the PCC experiment relies on successful crosslinking of LZD to its DNA binding sites, one would expect to see enhanced ligation between DNA fragments containing GCN4 sites, and consequently, increased crosslinking efficiencies of these fragments in the presence of the looping protein. However, the plots revealed two

Table 4.2: Cycle threshold values from all *in vitro* PCC qPCR products. norm – FPnorm+RPnorm primers, ctrl – FPctrl1+FPctrl2 primers

DNA	LZD73	CH_2O	A1+B2	A1+C2	A1+D2	norm	ctrl
CREB	-	-	29.91	32.17	34.86	16.65	28.02
CREB	-	+	28.55	31.22	34.78	16.75	27.34
CREB	+	-	28.70	30.51	37.25	16.61	27.16
CREB	+	+	28.80	30.10	35.43	16.58	27.11
IC-300	-	-	28.96	32.02	37.05	16.70	28.19
IC-300	-	+	28.75	31.98	36.35	17.16	28.09
IC-300	+	-	28.78	31.16	35.98	16.86	27.67
IC-300	+	+	28.69	30.41	36.98	16.85	27.54
IC-450	-	-	30.05	32.69	37.27	17.06	29.12
IC-450	-	+	29.03	31.87	36.94	16.63	28.01
IC-450	+	-	29.61	31.62	37.28	16.72	28.13
IC-450	+	+	29.33	30.22	36.64	16.90	27.67
IC-900	-	-	30.31	32.39	36.03	17.54	29.14
IC-900	-	+	29.87	32.28	35.94	17.16	28.69
IC-900	+	-	29.79	31.63	36.69	17.01	28.48
IC-900	+	+	29.24	30.25	34.70	17.13	27.89

things: (1) there are no changes in the crosslinking efficiencies in the region of interest on all samples in the absence of LZD73, and (2) there are no localized peaks in the crosslinking efficiencies on samples containing one GCN4 or two GCN4 sites 300 bp apart in the presence of LZD73. However, there is a very small peak in the crosslinking efficiency plot when the INV2 is either 450 or 900 bp away from CREB. This peak only shows up in samples containing LZD73. This looks promising in the sense that LZD73 has been previously shown to stabilize ∼450-bp DNA loops from cyclization experiments (Gowetski et al., 2013).

Taking into account the results from the attempt to enrich *in vitro* crosslinked complexes by affinity chromatography, and the qPCR, we have seen that the

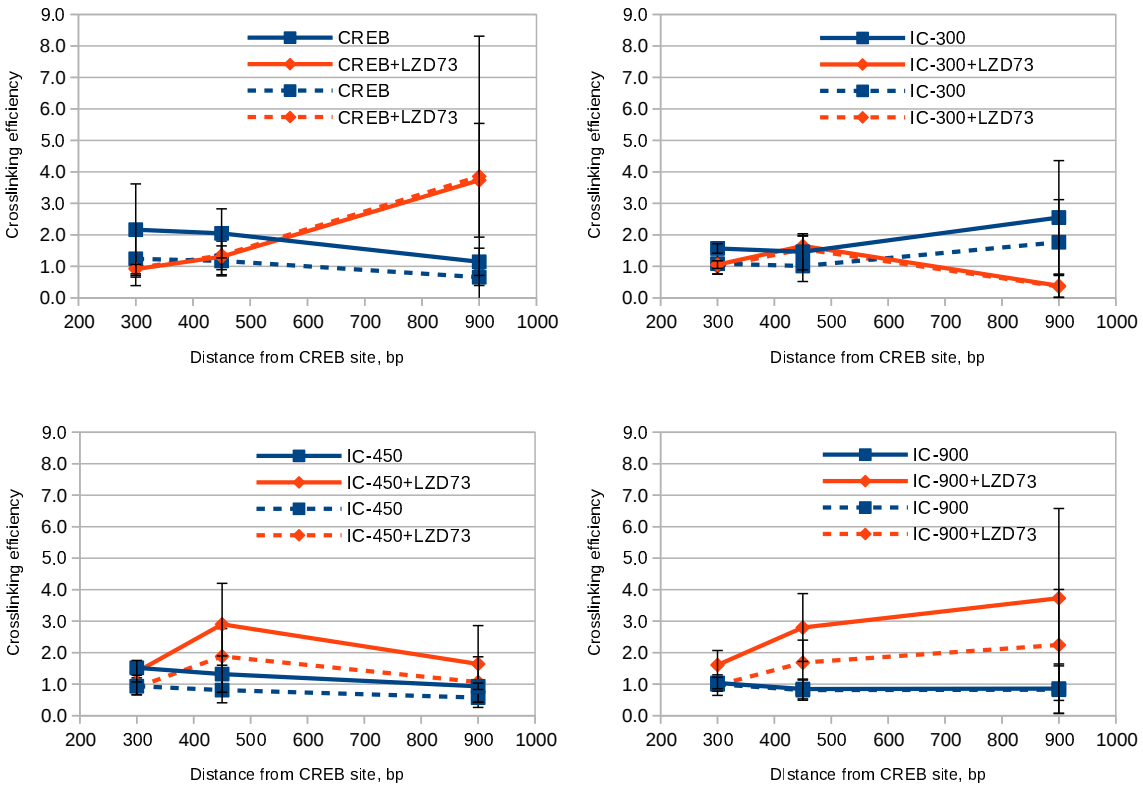


Figure 4.12: Crosslinking efficiencies among *in vitro* PCC templates, as a function of distance from the proximal CREB site. Solid plots are from data normalized to template yield, while dashed plots are from data normalized to the amount of intermolecular ligation.

His-tagged LZD proteins do not seem to form DNA crosslinks with formaldehyde. One key feature in formaldehyde-mediated protein-DNA crosslinks is to have nucleophilic residues from both the protein and the DNA that are close enough for the crosslinker to react with. Lu et al. (2010) elucidated structures of formaldehyde-induced crosslinks of four amino acids with three nucleotides, in which the Lys-dG crosslink gave the highest yield. A closer inspection of the LZD protein sequences (see Appendix 2.4) showed that our protein have a lot of lysines, but none are in the vicinity of the DNA binding region that makes crucial contacts with the DNA. We believe that the lack of nucleophilic amino acid residues in the DNA binding

domain is the main cause of the failure of our PCC experiments.

4.5 Conclusions

Results from both *in vivo* and *in vitro* PCC experiments showed that we have optimized crosslinking conditions that would give decent restriction enzyme and ligase activities. Given our current protocol, it seems that trace amounts of formaldehyde in our templates have considerably lowered the efficiency of digestion and religation, which affects the specific ligation product yields. Therefore, the crosslinking and template harvest methods have to be modified further to improve final product yields.

An important requirement for a successful 3C assay lies in the ability of the protein being studied to form specific crosslinks with its DNA binding site. However, the LZD protein sequences showed two lysine residues in the DNA binding domains, but neither of them are close enough to make crucial contacts with the DNA. In order to move the project forward, we need to generate DNA-crosslinkable LZD proteins. We have addressed this challenge by identifying four amino acids in the DNA binding domains, and mutating them into lysines. This will be discussed in more detail in the next chapter.

Chapter 5: Improving protein design

5.1 Objectives

Our initial reporter gene assay results indicate that 4har and lzee are fairly good repressors, but not efficient looping proteins on operators spaced 85.5 base pairs apart. Increasing the distance to 300-1150 failed to show that these proteins can loop DNA. There is a possibility that the C-terminal four helix bundle region does not form stably upon tetramerization, which would explain the non-formation of DNA loops with these proteins. Thus, the first part of this chapter discusses the initial attempt to improve the design to the LZT proteins, by designing a protein library and screening for mutants that can form more stable C-terminal four helix bundles and repress the reporter protein expression in our specialized *E. coli* strains.

The failure of the PCC mainly stems from the initial requirement that the protein being studied should be able to make crosslinks to DNA. Unfortunately, all our proteins don't have lysine residues located in the vicinity of their DNA binding sites that can make the crosslinks with DNA contacts. In the second part of this chapter, we discuss the lengths we went through in generating lysine mutants and selecting for better DNA crosslinkers, with the main goal of generating a version of the LZD proteins that can crosslink each end of the protein to its respective

GCN4 binding site. We hope that these crosslinkable proteins would work better in capturing *in vitro* and *in vivo* DNA loops by PCC.

5.2 Rationale

De novo protein design relies on the premise that the resulting protein folds into a structure that buries hydrophobic residues into the core of the protein and exposes the hydrophilic residues to the polar solvent. Moreover, there is a network of secondary structures usually stabilized by hydrogen bonds that keeps the protein folded. With these in mind, one can design a particular secondary structure by choosing alternating polar and nonpolar amino acid sequences with a periodicity that matches the desired secondary structure (Kamtekar, Schiffer, Xiong, Babik, & Hecht, 1993). In the case of an alpha-helical protein, the pattern of polar and nonpolar amino acids should be close to the 3.6-residue repeat. Given this pattern, a coiled-coil motif can be generated if the first and fourth amino acids in the heptad repeat would either be a nonpolar amino acid (usually valine or leucine), and the rest are polar residues.

The LZT protein design incorporates both parallel coiled-coil and the antiparallel four helix bundle motifs. In order to transition from one fold into another while maintaining alpha helical registry, linker sequences are designed to allow the parallel alpha helices to open up a little bit to allow the kitty positioning of the antiparallel alpha helices. Thus, we designed the LZT library by optimizing sequence patterns in this linker region. Figure 5.1 illustrates the strategy of building the LZT protein

library.

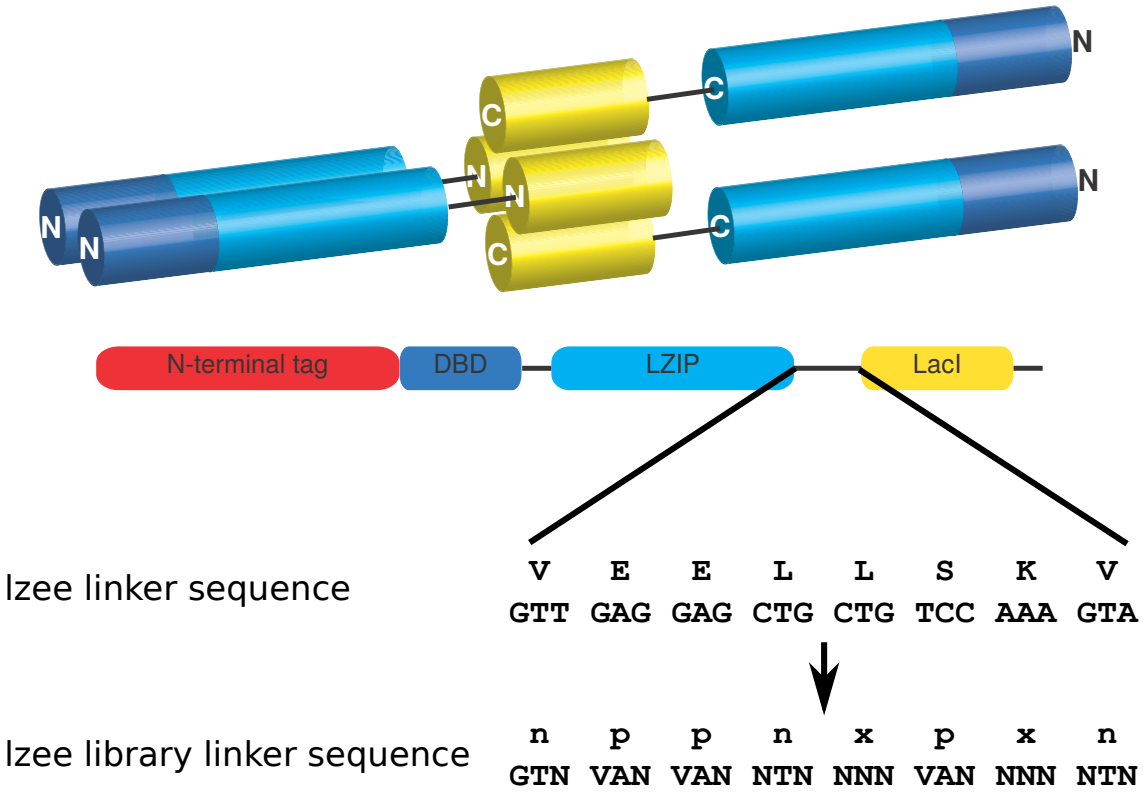


Figure 5.1: Strategy for designing the LZT protein library, based on the binary patterning of amino acid sequences (Kamtekar et al., 1993). We used lzee as the protein template to build the library on. To allow seamless transition into the antiparallel coiled-coils, nonspecific amino acids (x) are placed in the e and g positions of the heptad repeat. protein sequences: n – nonpolar; p – polar; x – any amino acid. DNA sequences: V – A, G or C; N – A, G, C, or T

The challenge in generating DNA-crosslinkable proteins is to pick residues that can make formaldehyde adducts with DNA without disrupting the existing network of bonds between the protein and the its DNA binding site. In order to do this, we first identified amino acids in the DNA binding domain of the LZD proteins by picking out residues in the crystal structure of the GCN4 bZip domain (PDB file 1DGC) that are within 2 – 3 Å from the 6-amino group of adenine or 2-amino group of guanine.

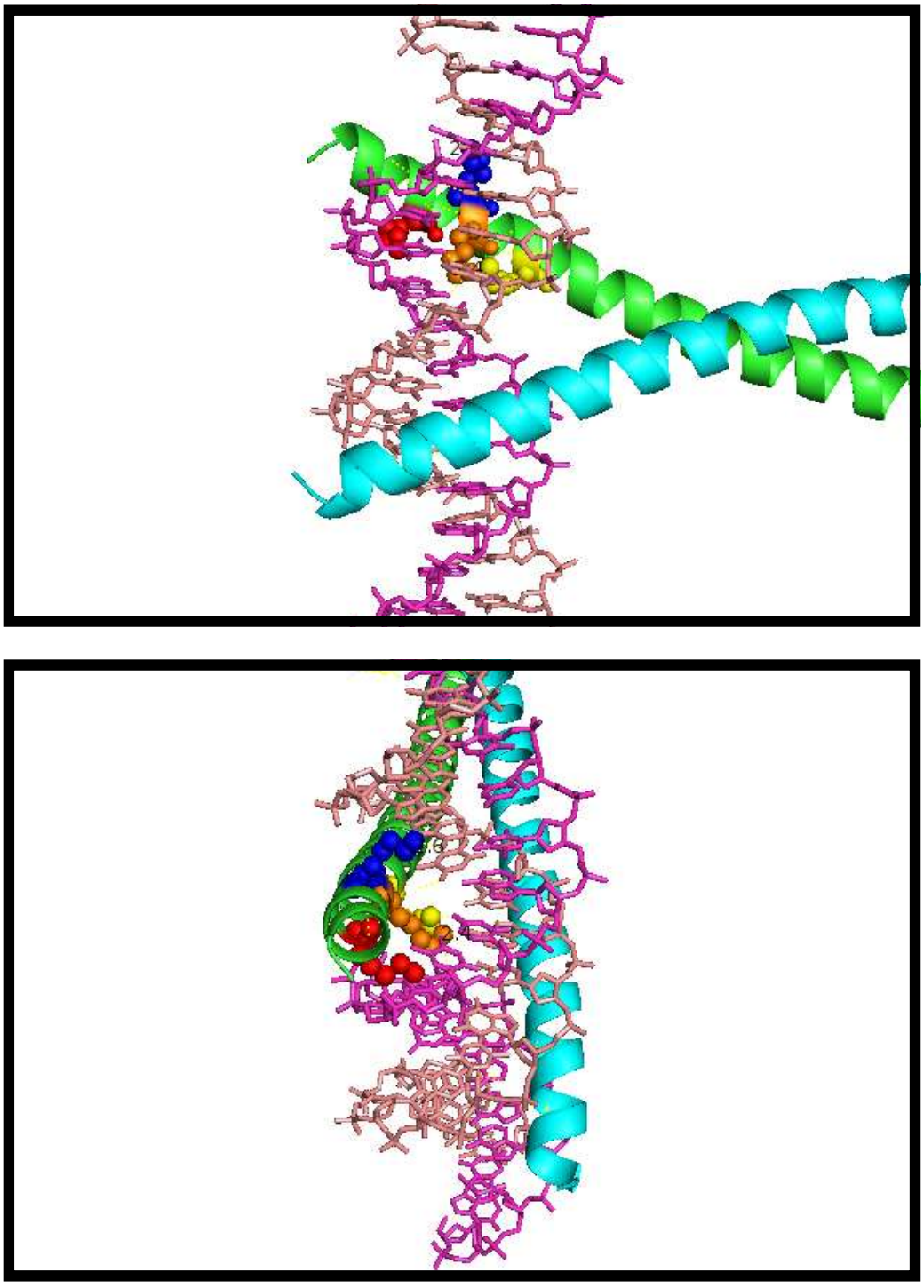


Figure 5.2: Two views of the LZD87 mutants, highlighting the locations of the amino acids mutated into lysine residues. red spheres: T6K, blue spheres: A8K, orange spheres: A9K, yellow spheres: R13K. The images were generated using PyMOL, using PDB file 2DGC as template.

Each of the these amino acids are then mutated into lysine, with the idea that lysine would be in close proximity to these bases to be crosslinked with formaldehyde. The same amino acids in the C-terminal DNA binding domain are mutated, which would generate our first set of LZD single mutants. Then, each of the single mutants would be tested for its crosslinking ability; once we have identified better crosslinkers from each end of the protein, then we can generate a double mutant that can crosslink DNA from both ends of the protein.

5.3 Materials and Methods

Oligonucleotides used for cloning were purchased from IDT, while molecular biology reagents and enzymes were purchased from NEB. All other reagents were purchased from Fisher Scientific, Sigma, and USB Corporation. PCR was done on either Eppendorf Mastercycler or MJ Research PTC-200 thermal cycler. Protein purifications were done on the ÅKTA FPLC instrument (Amersham Biosciences), attached with a 1-mL prepacked HisTrap HP (GE Healthcare) column previously charged with 0.2 M cobalt acetate.

5.3.1 Generating the LZT protein library

The parent plasmid for cloning in the LZT library was prepared as follows: Mutagenesis PCRs were done on lacUV5-lzee/pGEX6P1 plasmid using the primer pairs that (1) introduced an upstream *Spe*I site in the linker region, and (2) that introduced a downstream *Mfe*I site.

1. **lzeeFP-BamHI:** 5'-CCCCTGGGATCCGACCCAGCGGCACTG-3'

   mutRP-SpeI: 5'-GCCAGAGCACGTACTTTGGACAGCAGCTCCTCAA CTAGTTTTTTCAG-3'

2. **mutFP-SpeI-MfeI:** 5'-CGTCTGAAAAAACTAGTTGAGGAGCTGCTG TCCAAAGTACGTGCTCTGGCTGATTCTCTGATGCAATTGGCTCGC C-3'

   lzeeRP-EcoRI: 5'-CGACCCGGGAATTCTACTACTGACCGGATTCCA GACGG-3'

The first reaction generated a 203 bp product, while the second reaction gave a 121 bp product. Each PCR product was purified using the PCR Purification Kit (Thermo Scientific), and then mixed together for mutually primed synthesis of the complete lzee gene insert that contains both *Spe*I and *Mfe*I sites in the linker region (274 bp). The resulting product was digested with *Bam*HI and *Eco*RI and cloned into the *Bam*HI – *Eco*RI digest of lacUV5-pGEX6P1 plasmid. The recombinant plasmid was then transformed into XL1-Blue cells and plated onto LB amp selection plates. Correct transformants were identified by *Spe*I or *Mfe*I digestion and DNA sequencing. This plasmid was then identified as lacUV5-lzeeV2/pGEX6P1.

To generate the LZT library, the following overlapping oligonucleotides were annealed together and amplified using Klenow fragment:

- **LZEElib-SpeI-v2:** 5'-CTGAAAAAACTAGTNVANVANNTNNNNVANN NNNTNCGTGCTCTGGCT-3'

- **LZEElib-MfeI-v2:** 3'-GCACGAGACCGACTAAGAGACTACGTTAACC

GAGCGGTC-5'

The resulting Klenow product was digested with *Spe*I and *Mfe*I, and the 50-bp digestion product was isolated by native PAGE and worked up by phenol/chloroform extraction and ethanol precipitation. The library was then cloned into the *Spe*I – *Mfe*I digest of lacUV5-lzeeV2/pGEX6P1. The ligation product is purified and worked up by phenol/chloroform extraction and ethanol precipitation.

As a positive control for library insertion, the following oligonucleotide was substituted for LZEElib-SpeI-v2 in the first step: LZEElibctrl-SpeI 5'-CTGAAAAA ACTAGTTGAGGAGCTCCTGTCCAAGCTTCGTGCTCTGGCT-3'. Successful insertion of this sequence into the plasmid is verified by *Sac*I or *Hin*dIII digestion.

5.3.2 Generating the LZD87 lysine mutants

The single mutants on the N-terminal DNA binding domains were generated by site-directed mutagenesis PCR using LZD87/pRSETA as the plasmid template and the following phosphorylated primers:

- reverse primer **NtermDBDmutRP:** 3'-CCTAGGCTAGGTCGACGAGAC TTCGCTCGAGCC-5'
- forward primers
  - **T6KmutFP:** 5'-AACAAGGAAGCTGCTCGACGGAGCCGAGCTCG GAAGCTGCAACGAATG-3'
  - **A8KmutFP:** 5'-AACACTGAAAAAGCTCGACGGAGCCGAGCTCG GAAGCTGCAACGAATG-3'

- **A9KmutFP:** 5'-AACACTGAAGCTAAACGACGGAGCCGAGCTCG GAAGCTGCAACGAATG-3'
- **R13KmutFP:** 5'-AACACTGAAGCTGCTCGACGGAGCAAGGCTC GGAAGCTGCAACGAATG-3'

On the other hand, the single mutants on the C-terminal DNA binding domains were made similarly using the following phosphorylated primers:

- reverse primer **CtermDBDmutRP:** 3'-CTTGACGTCTTCAATGTCGCC CACTTCGCTCGAGCC-5'
- forward primers
  - **T77KmutFP:** 5'-AACAAGGAAGCTGCTCGACGGAGCCGAGCTC GAAAGGCTGCTCTGAAG-3'
  - **A79KmutFP:** 5'-AACACTGAAAAAGCTCGACGGAGCCGAGCTC GAAAGGCTGCTCTGAAG-3'
  - **A80KmutFP:** 5'-AACACTGAAGCTAAACGACGGAGCCGAGCTC GAAAGGCTGCTCTGAAG-3'
  - **R84KmutFP:** 5'-AACACTGAAGCTGCTCGACGGAGCAAGGCTC GAAAGGCTGCTCTGAAG-3'

The ends of the resulting linear mutated plasmid products were blunt ligated together and cloned into XL1-Blue cells for plasmid propagation. Correct transformants were identified by DNA sequencing.

5.3.3 *In vivo* crosslinking assays on LZD87 single mutants

In vivo crosslinking for the LZD87 mutants was done similarly to the crosslinking method for PCC, with a few modifications. BL21-AI cells were transformed with plasmids containing the GCN4 binding sites (IC-450) and the expression plasmid for the LZD87 single mutants. The resulting transformants were plated on LB selection plates containing amp, kan and cam. Surviving colonies were then grown in liquid cultures and the LZD proteins were induced with 0.2 % arabinose for 15 hours at 37 °C with agitation. The cells were harvested and resuspended in 1x PBS buffer ± 1 % formaldehyde. The crosslinking step was done at 37 °C for 10 minutes, after which, the reactions were quenched with 350 mM glycine. The crosslinked cells were washed once with 1x PBS to remove residual formaldehyde, and then lysed with BugBuster® reagent in Native Binding Buffer (10 mM MES pH 6, 500 mM NaCl, 20 mM imidazole) at ambient temperature for 20 minutes. The crosslinked His-tagged proteins were isolated from the lysate using batch purification with Ni Sepharose FF beads (GE Healthcare). The crosslinks were removed from the supernatant after mixing with the affinity beads and the initial eluate by incubating the samples at 70 °C for at least 8 hours. prior to DNA workup using the PCR purification kit. Finally, the worked up samples were analyzed for IC-450 yields using PCR.

5.4 Results and Discussion

5.4.1 Screening the LZT library for better repressors

We used the *in vivo* repression assay coupled with a colorimetric screen in Chapter 3 to look for better repressors in the LZT library. First, the ligation mixture generated in Section 5.3.1 was transformed into FW102/CW-237.5, a reporter strain containing a distal CREB and a proximal wild-type GCN4 site 237.5 bp apart. The transformants are then plated onto tetrazolium-glucose and tetrazolium-lactose plates containing amp, kan, and strep. Pink colonies from both colorimetric plates were then selected for the repression assays.

Repression assay results on selected library transformants showed that no transformant stood out in terms of its ability to significantly repress the reporter protein (see Figure 5.3). The result for transformant MP73-33 (top plot) was eventually ruled out as an outlier when the assay was repeated (middle plot). Moreover, sequencing results for MP73-33 showed a truncated sequence that only included the bZip sequence. From the transformants that were analyzed, we verified the gene insert sequences for eight of them. Three of them showed the complete sequence that matches the intended amino acid pattern (MP113-1, MP113-5, MP113-7), two transformants did not clone in the library insert but got translated up to the four-helix bundle sequence (MP95-15, MP95-25), two had truncated sequences (MP73-33 and MP73-49), and one had a doubly-inserted library sequence (MP113-10).

There are two possible outcomes for the transformants growing on tetra-lac

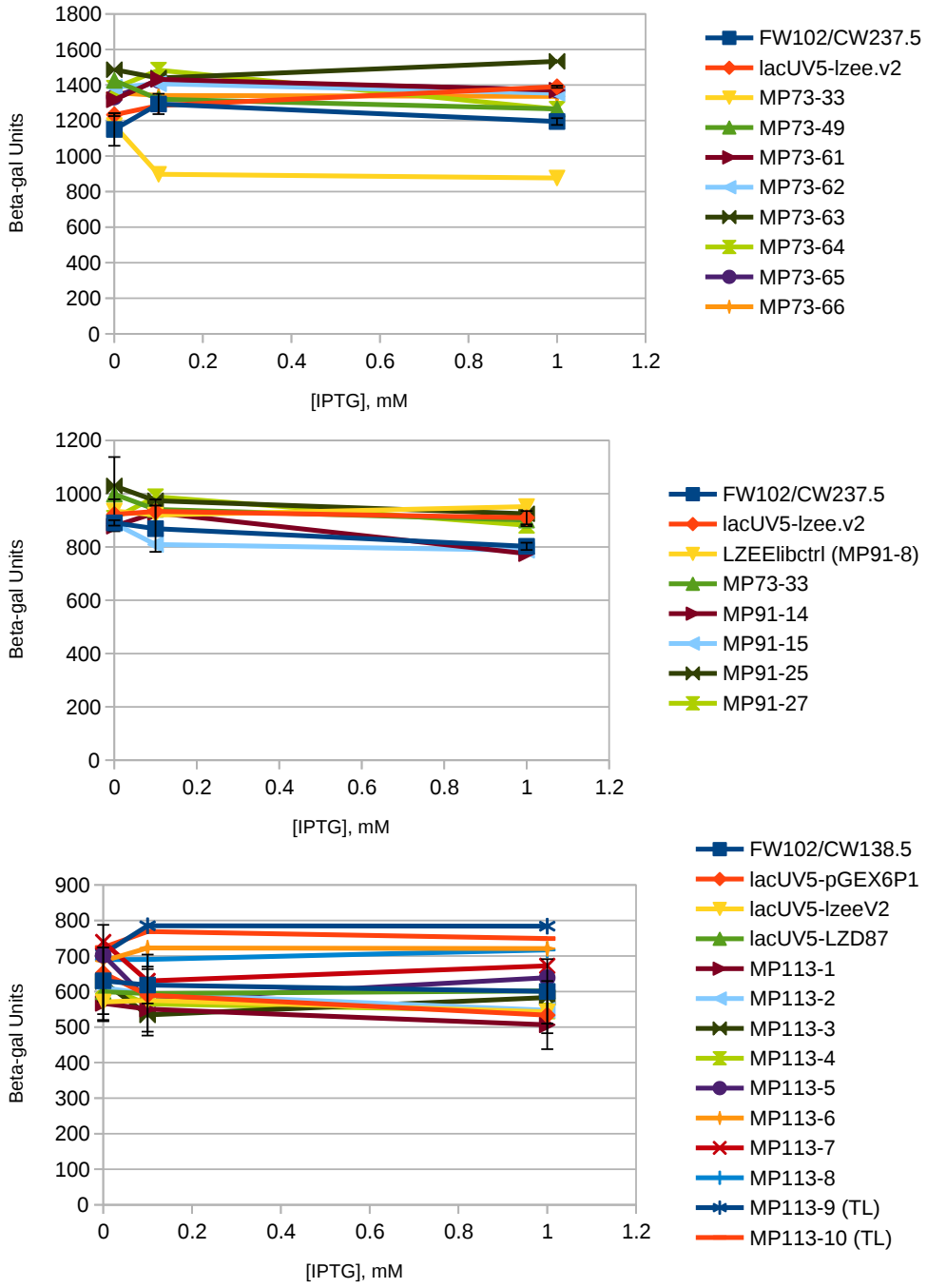


Figure 5.3: Screening the LZT library for better repressors. Transformants MP113-9 and MP113-10 were selected from tetrazolium lactose plates, while the rest were selected from tetrazolium glucose. None of the transformants analyzed showed significant repression of the reporter protein.

plates. If the mutant is a very efficient repressor as lzee, it will not grow in tetra-lac media. Otherwise, it will either be white or pink. Given the proposed recombination events discussed in Chapter 3.4.8 (page 80), it looks that what we are actually analyzing in the repression assays are the weak repressors.

5.4.2 *In vivo* crosslinking results

Next, we tested four of the LZD87 single mutants in their ability to crosslink DNA *in vivo*. Figure 5.4 shows that all of the LZD87 single mutants but T6K-LZD87 analyzed seem to have good expression amounts. PCR results from both the supernatants post-mixing with the affinity beads and the eluates showed comparable yields of IC-450 plasmid (see Figures 5.5 and 5.6). This could mean two things: (1) the template concentrations have not been optimized to linear amounts, and (2) the LZD87 mutant may not have crosslinked successfully to the DNA.

At this point, we could have proceeded to quantifying the plasmid yields by qPCR, but we decided to do the crosslinking *in vitro* and see if we could get clearer results.

5.4.3 *In vitro* crosslinking results

In the previous chapter, we have observed that formaldehyde had significantly affected the restriction enzyme and ligation activities in our PCC experiments. Thus, we first tested several methods of quenching and removing residual formaldehyde in crosslinked reactions containing DNA to determine the best way to remove

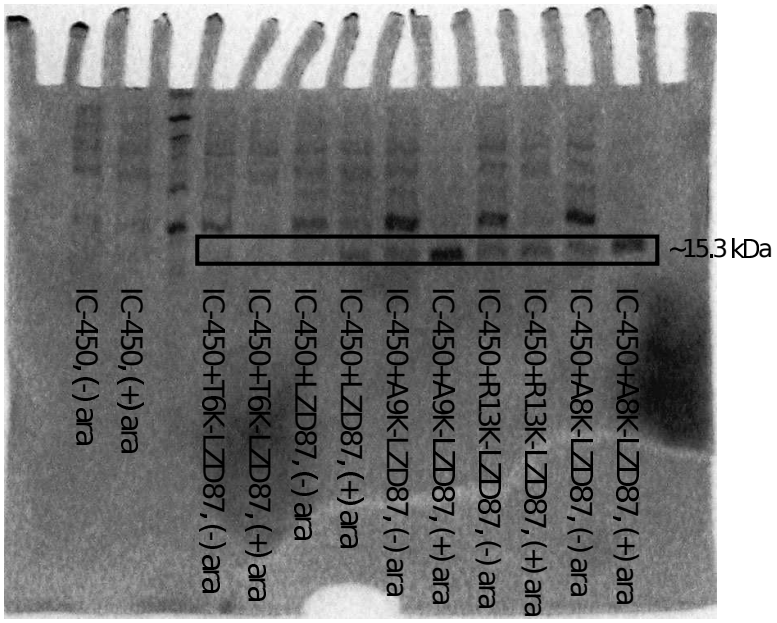


Figure 5.4: Protein expression on LZD87 single mutants. SDS-PAGE of crude lysates of BL21-AI transformed with LZD87/pRSETA plasmids, 15 hours post-induction 0.2 % arabinose. Results showed good expression levels on all LZD87 single mutants.

excess formaldehyde in our *in vitro* PCC samples. Figure 5.7 showed that removing formaldehyde by isopropanol precipitation was indeed more efficient than simple quenching of the excess formaldehyde with glycine, Tris or the restriction enzyme buffer.

Then, we adapted a ChIP-like method to demonstrate a quick and simple way for the LZD87 mutants to crosslink DNA *in vitro*. Crosslinking the protein to the plasmid DNA coupled with a purification step to isolate the His-tagged protein allows us to concentrate the amount of DNA that is actually bound to the protein. This way, we are certain that subsequent enzymatic treatments (i.e. digestion and religation for PCC) are done on samples that may be forming DNA loops.

A binding reaction containing 5 nM IC-450 plasmid and 25 nM purified LZD87

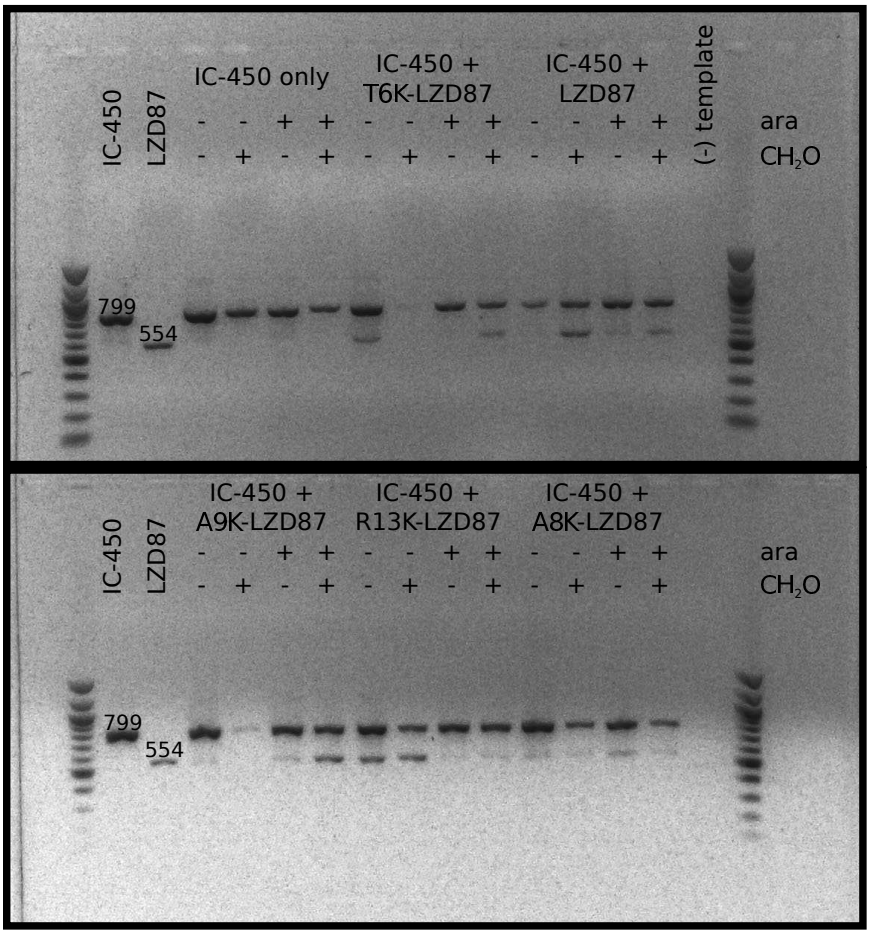


Figure 5.5: Plasmid yields from batch purification of crosslinked protein-DNA complexes. Agarose gel electrophoresis of PCR products from reverse-crosslinked and worked up initial supernatants from batch purification of His-tagged LZD87 in BL21-AI, using primers that amplify the GCN4 sites in IC-450 (799 bp PCR product) and the gene inserts in the recombinant pRSET plasmids (554 bp PCR product).

mutant was prepared in H/L buffer (20 mM HEPES pH 7.7, 2 mM Na_2(ATP), 0.2 % glycerol, 0.01 % IGEPAL, 4 mM $MgCl_2$, 10 mM DTT, 100 μg/mL BSA) and incubate at ambient temperature for 30 minutes. Then, the solution was crosslinked with 0.1 % formaldehyde at 37 °C for 10 minutes. The complexes were isolated by isopropanol precipitation, and resuspended in restriction enzyme buffer containing the restriction enzyme. We used *Apo*I as the restriction buffer (IC-450 plasmid has 9

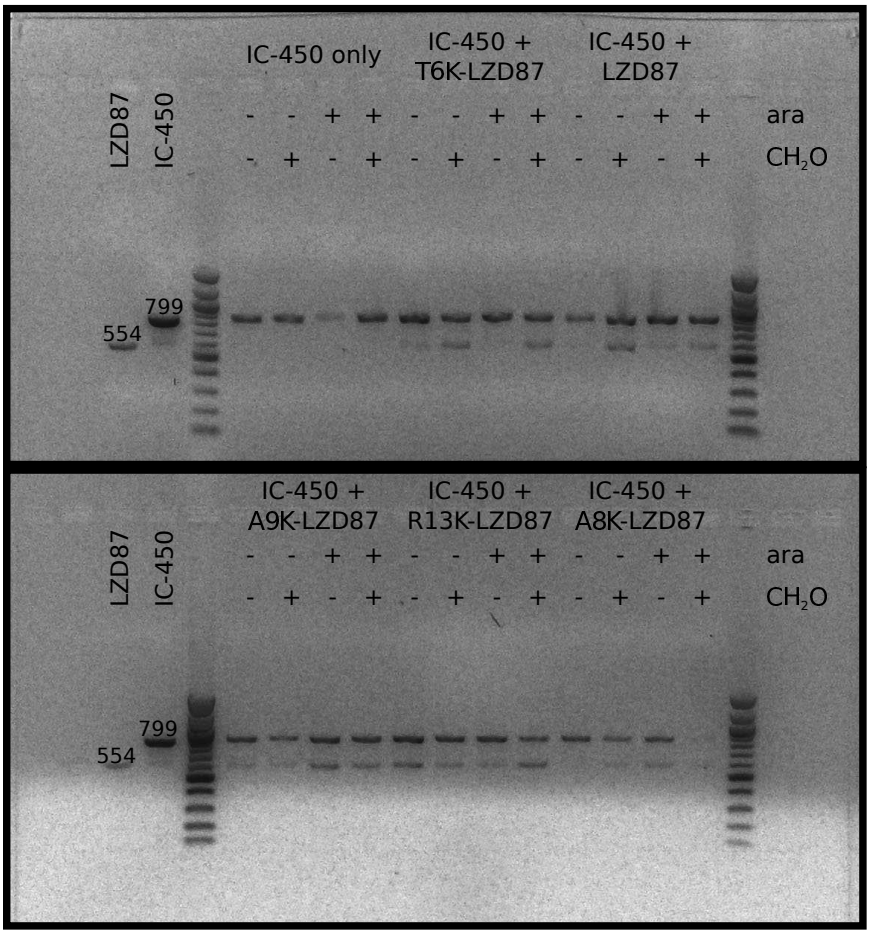


Figure 5.6: Plasmid yields from batch purification of crosslinked protein-DNA complexes. Agarose gel electrophoresis of PCR products from reverse-crosslinked and worked up initial eluates from batch purification of His-tagged LZD87 in BL21-AI, using primers that amplify the GCN4 sites in IC-450 (799 PCR product) and the gene inserts in the recombinant pRSET plasmids (554 bp PCR product).

*Apo*I cut sites, with a 272-bp fragment containing CREB site, and a 732-bp fragment containing the INV2 site). The mixture was digested for 1 hour, and the digestion products were subsequently labeled with α^{32}P ATP. The labeled samples were then diluted in Native Binding Buffer and passed through His SpinTrap columns to isolate His-tagged proteins, which may or may not be attached to the DNA. Radioactivities in the initial flowthroughs and the eluates were measured by scintillation counting

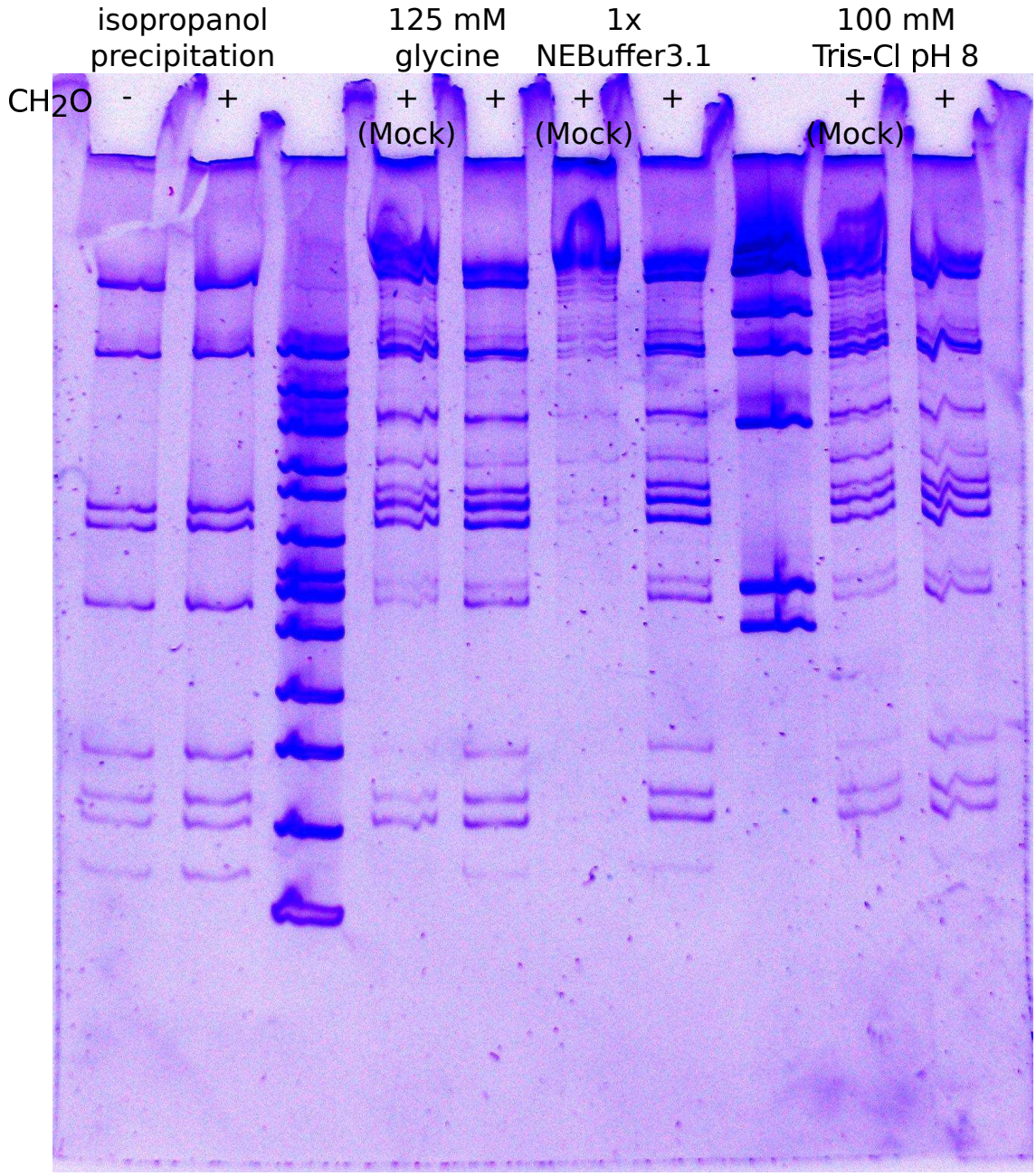


Figure 5.7: Comparing various methods of removing formaldehyde in crosslinked DNA. Native PAGE of *Apo*I digestion products of IC-450 ± CH_2O. 2.5 nM IC-450 plasmid was incubated in H/L buffer ± 0.1 % formaldehyde at 37 °C for 10 minutes. Crosslinked samples were quenched in the stop solutions at 37 °C for 1 hour before digesting with *Apo*I (0.1 U/μL). Mock – IC-450 was only added after the quenching step.

and the samples were analysed by SDS-PAGE.

SDS-PAGE analysis showed most of the *Apo*I digestion products ended up in the initial flowthrough, which indicated that most of the DNA was not bound and/or crosslinked to the protein (see Figure 5.8). However, no bands were detected in the initial eluates (gel picture not shown), which showed that none of the LZD mutants tested so far can crosslink to its DNA binding site. It is possible that the formaldehyde may not be an efficient crosslinker in our system.

Finally, we repeated the experiment on the LZD87 double mutants with UV crosslinking using a 254 nm UV light with a 120000 μJ/cm^2 energy output. SDS-PAGE analysis of all batch purification samples showed similar results as that of the formaldehyde crosslinking experiment.

5.5 Conclusion

Based on the results presented, we have not identified a better LZT repressor (or possibly a better looping protein), nor we have seen clear evidence of DNA crosslinking in any of the LZD87 mutants that we have tested.

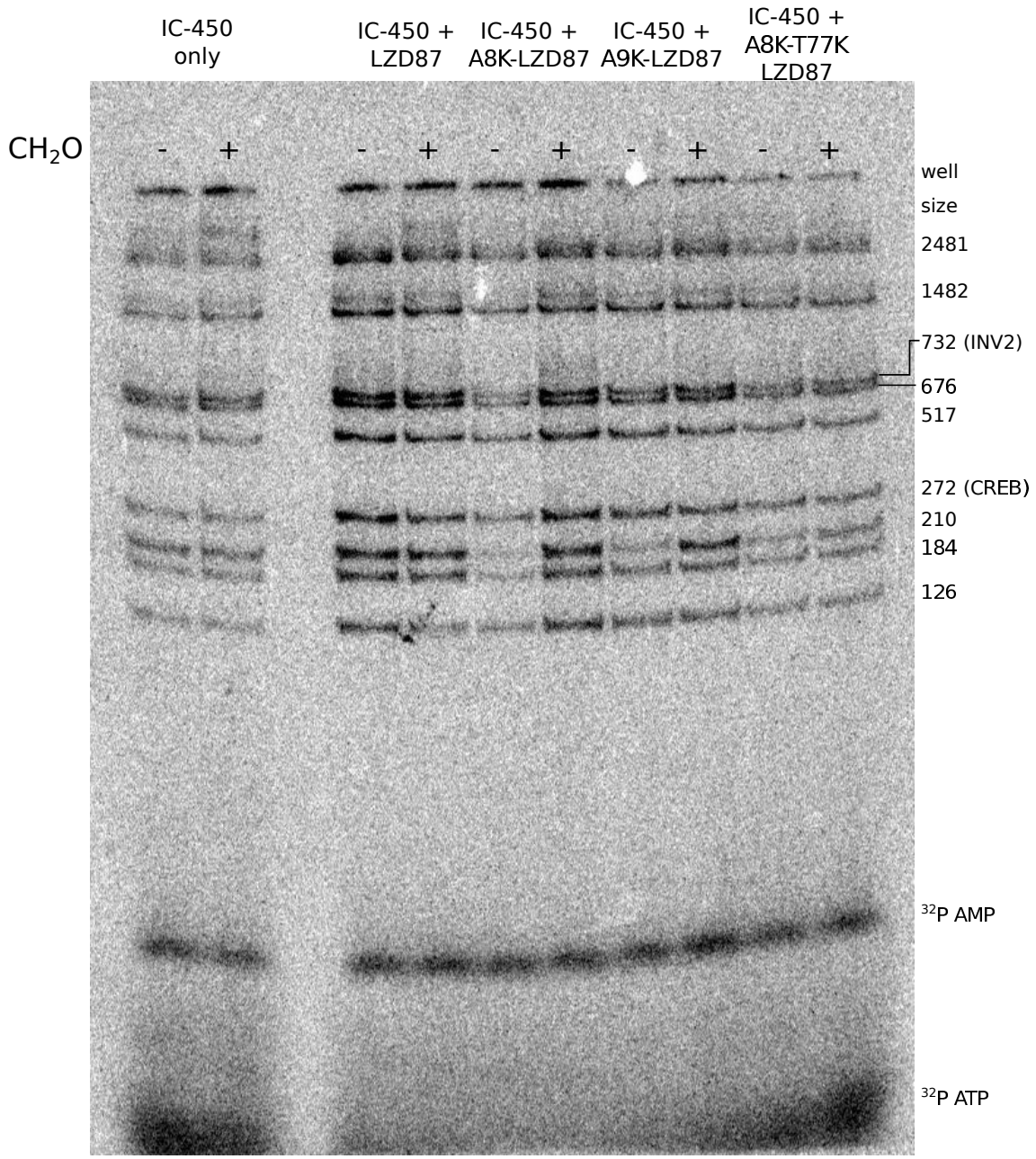


Figure 5.8: Phosphorimage of *Apo*I digestion products of formaldehyde-crosslinked IC-450 ± LZD87 mutants after batch purification, initial supernatants. Bands were visualized in 10 % (40:1) polyacrylamide gel in Tris-glycine-SDS buffer.

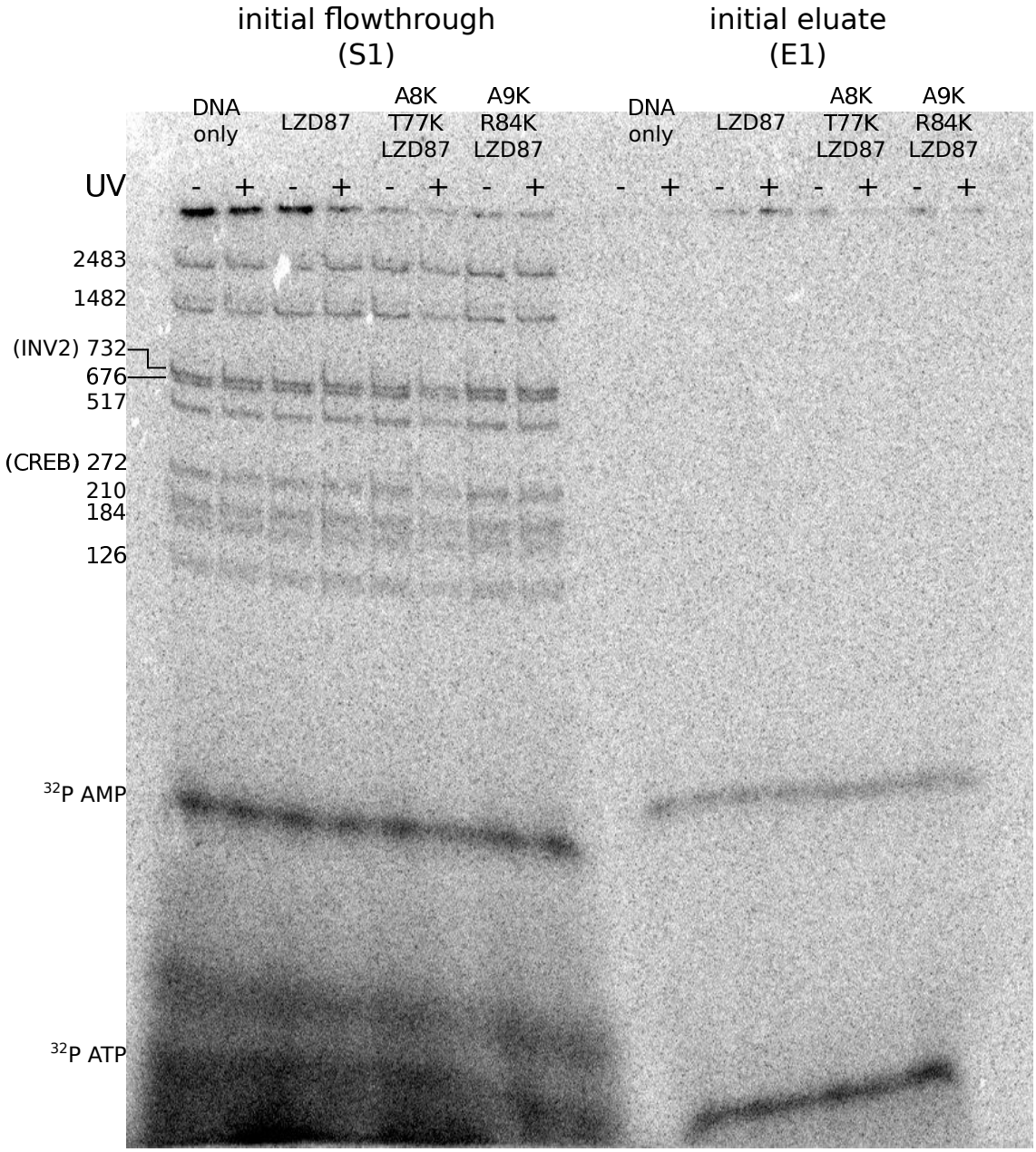


Figure 5.9: Phosphorimage of *Apo*I digestion products of UV-crosslinked IC-450 ± LZD87 mutants after batch purification. Bands were visualized in 10 % (40:1) polyacrylamide gel in Tris-glycine-SDS buffer.

Chapter 6: Conclusions and Future Directions

In Chapter 2, we have demonstrated that none of the tetrameric DNA binding proteins can form a protein-DNA sandwich complex, which is a necessary precursor to illustrate DNA looping. In Chapter 3, we have seen relatively weak cooperative repression in reporter strains containing high-affinity GCN4 operators >300 bp apart upon dose-dependent expression of lzee and LZD87, although there is no clear evidence whether the repression is due to DNA looping. Moreover, we have also seen host either cell toxicity or significantly low expression of reporter protein upon induction of lzee and LZD87. In Chapter 4, we have seen a weak interaction between DNA loci containing GCN4 sites spaced 450 bp, and possibly 900 bp, apart in the presence of LZD73. Finally, in Chapter 5, we were not successful in selecting a better LZT repressor, or have showed that one of the LZD87 mutants can make crosslinks to the DNA.

What have we learned from all these experiments? One thing stands out – it is more difficult than initially thought to make DNA loops *in vivo* using very small and relatively rigid DNA looping proteins. On one hand, it is possible that in forming small loops, the energy cost in bending the intervening DNA to conform to the relatively inflexible looping protein is too great such that it is energetically more

favorable for the host cell to inactivate either the protein, by introducing mutations to the expression plasmid, or the reporter gene. On the other hand, in large loop formation, partially unwinding the supercoiled DNA, or to bending the looping protein in order to wedge itself within the supercoiled DNA, so as to accommodate binding to two operators with the correct spatial orientation may also be too large. The results presented here support the thermodynamic model that DNA flexibility plays a more passive role in *in vivo* DNA loop formation.

6.1 Characterization experiments on the LZD proteins

We can still evaluate the *in vivo* looping efficiency of the LZD proteins using our repression assay, given the following modifications. First, using the pGEX plasmids as the expression plasmids for our repression assay was rife with problems. Using expression plasmids designed for protein overexpression are not suitable for looping assays where the desired response should be sensitive to increased local concentration of the protein at the primary operator via cooperativity at a distance, and not by protein saturation at the primary site. At the same time, having too many homologous sequences between the expression plasmid and the F' episome, which would most likely result to different recombination events, would complicate our analysis. Thus, we suggest to clone the LZD genes into the expression plasmid Whipple developed to complement the reporter strains in his genetic assay.

Second, we can further optimize the combination of weak proximal and strong distal operators in our looping strains, specifically for characterizing the LZD pro-

teins. By design, the N-terminal end of the protein binds with high affinity to CREB, while the C-terminal end binds very well to INV2. However, we could not tell for sure which end of the protein actually binds to which operator during the repression assay. Alternatively, we can optimize amino acid sequences in the DNA binding domains of LZD proteins to engineer ends with low or high binding affinities to the operators.

Third, we can introduce a single flexible hinge in the extended LZIP region in the LZD proteins by introducing prolines or glycines. These amino acid residues are known to break protein secondary structures due to steric hindrance of the side chain or high conformational flexibility, respectively. This would allow some degree of protein flexibility, which may be able to stabilize smaller DNA loops *in vivo*. If the new proteins can stabilize smaller loops than those of the LZD proteins, this would further support the thermodynamic model that looping protein flexibility drives *in vivo* DNA looping.

6.2 Modifications on the LZT protein library

From a design standpoint, switching from a parallel coiled-coil to an antiparallel four helix bundle requires opening up the C-terminal end of the coiled-coil to accommodate the alpha helices from the other coiled-coil. A simple adjustment to the protein library sequence is to introduce one or two glycine residues in the heptad linker sequence. This would introduce breaks in the parallel coiled-coils, which would aid in the correct folding of the four helix bundle at the C-terminal ends.

However, this might introduce some degree of flexibility to the whole protein, which may contribute to the overall stability of DNA loops.

The protein library can then be probed using the repression assay as currently designed, to select for better looping proteins. The selected mutants can then be expressed in large-scale quantities and characterized further by ultracentrifugation techniques to verify if they can form tetramers in solution.

6.3 Conclusions

We have utilized and optimized several molecular tools in characterizing the *in vitro* and *in vivo* looping abilities of artificially designed DNA binding proteins. The failure of the LZD proteins to stabilize *in vivo* DNA loops further supports the idea that inherent protein flexibility is a necessary requirement in the formation of DNA loops. The inability of the putatively tetrameric lzee protein to form *in vitro* sandwich complexes with DNA, yet functions as a highly effective repressor protein that triggers host cell toxicity or permanent repression, can be utilized in designing other genetic assays that detect recombination, which could be a consequence of DNA looping. This work embodies an additional step in developing an arsenal of tools, not just in understanding the thermodynamics of DNA loop formation, but also provide us tools to be used to modulate *in vivo* gene expression at will.

Appendix 1: Growth Media and Selection Plates for Microbiology Work

1.1 General Directions

- Use autoclaved glasswares or sterile containers for storing solutions and media.
- Filter all solutions through a 0.2-micron filter, or autoclave at 121 °C for at least 20 minutes. Cool autoclaved growth media to ∼55 °C before mixing in the antibiotics.
- Store solutions at room temperature, growth media and agar plates at 4 °C.

1.2 Solutions

- 1 M $MgCl_2$, 50 mL (store in 10-mL aliquots)
- 1 M $MgSO_4$, 50 mL (store in 10-mL aliquots)
- 2 M glucose, 50 mL (store in 10-mL aliquots)
- glycerol, 50 mL (autoclave for 20 minutes)

1.3 Growth Media

1. Luria-Bertani (LB): 1% tryptone, 0.5% yeast extract, 1% NaCl

   Dissolve 10 g tryptone, 5 g yeast extract, and 10 g NaCl in deionized water to make 1 L solution.

   (a) LBAG (for preparing glycerol stock solutions, and working cultures of pGEX-transformed BL21 cells): LB + 100 μg/mL amp, 20 mM glucose

   Add 1 mL 100 mg/mL ampicillin and 10 mL 2 M glucose per 1 L autoclaved LB media.

   (b) LB amp tet (for preparing plasmid midipreps, minipreps or glycerol stock solutions of XL1-Blue cells transformed with pGEX or pRSET): LB + 100 μg/mL amp, 30 μg/mL tet

   Add 1 mL each of 100 mg/mL ampicillin and 30 mg/mL tetracycline per 1 L autoclaved LB media

(c) LB kan cam tet (for preparing plasmid midipreps, minipreps or glycerol stock solutions of XL1-Blue cells transformed with promoter-operator constructs): LB + 50 μg/mL kan, 50 μg/mL cam, 30 μg/mL tet

Add 1 mL each 50 mg/mL kanamycin, 50 mg/mL chloramphenicol and 30 mg/mL tetracycline per 1 L autoclaved LB media

(d) LB kan strep (for preparing working cultures of reporter strains containing intact *lacZYA* in the F' episome): LB + 50 μg/mL kan, 100 μg/ml strep

Add 1 mL each 50 mg/mL kanamycin, and 100 mg/mL streptomycin per 1 L autoclaved LB media

(e) LB amp kan strep (for preparing working cultures of reporter strains containing pGEX (or pBAD)): LB + 100 μg/mL amp, 50 μg/mL kan, 100 μg/ml strep

Add 1 mL each 100 mg/mL ampicillin, 50 mg/mL kanamycin, and 100 mg/mL streptomycin per 1 L autoclaved LB media

2. SOB: 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$

   Dissolve 20 g tryptone, 5 g yeast extract, g NaCl, g KCl in deionized water to make 1 L solution. Add 10 mL each of 1 M $MgCl_2$ and $MgSO_4$ per 1 L autoclaved media just before using.

3. SOC (for cell transformations, *should be freshly prepared*): SOB + 20 mM glucose

   Add 10 mL 2 M glucose per 1 L autoclaved SOB media

4. 2X YTA: 1.6% tryptone, 1% yeast extract, 0.5% NaCl, 100 μg/mL amp

   Dissolve 16 g tryptone, 10 g yeast extract, and 5 g NaCl in deionized water to make 1 L solution. Add 1 mL 100 mg/mL ampicillin to autoclaved media.

1.4 Agar Plates

- Add 15 g agar per 1 L growth media before autoclaving.
- Keep the working area clean and sterile by working near a bunsen burner.
- Cool down media to 55 °C before mixing in the antibiotics. Pipet 20 mL per sterile Petri dish. Let the solution set at room temperature before storing the plates upside down, wrapped in aluminum foil.

1. LBAG – for selecting BL21 cells containing intact pGEX after cell transformation

2. LB amp tet – for selecting XL1-Blue cells containing either intact pGEX, pRSET, or pBAD after cell transformation

3. LB kan cam tet X-gal (or Bluo-Gal) ± IPTG (or arabinose) – for selecting XL1-Blue cells containing intact promoter-operator constructs after cell transformation

4. LB kan strep X-gal (or Bluo-gal) ± IPTG (or arabinose) – for selecting reporter strains containing intact *lacZYA* in the F' episome

5. LB amp kan strep X-gal (or Bluo-gal) ± IPTG (or arabinose) – for selecting reporter strains containing pGEX (or pBAD) after cell transformation

6. MacConkey: 1.7% pancreatic digest of gelatin, 0.3% peptones (meat and casein), 1% lactose, 0.15% bile salts #3, 0.5% NaCl, 1.35% agar, 0.003% neutral red, 0.0001% crystal violet

   Dissolve 50 g MacConkey powder mix in deionized water to make 1 L solution. *Boil for 1 minute to completely dissolve the powder before autoclaving.* Mix in the required antibiotics after the solution has been autoclaved and cooled to 55 °C before pouring into Petri dishes.

7. Tetrazolium: Antibiotic Medium 2 (0.15% beef extract, 0.3% yeast extract, 0.6% peptone, 1.5% agar), 0.005% tetrazolium chloride

   Dissolve 25.5 g Antibiotic Medium 2 powder mix and 50 mg 2,3,5-triphenyl-2H-tetrazolium chloride in deionized water to make 1 L solution. *Boil for 1 minute to completely dissolve the mixture before autoclaving.* Mix in 50 mL 20% glucose, 50 mL 20% lactose, or 10 mL 20% arabinose, and the required antibiotics after the solution has been autoclaved and cooled to 55 °C before pouring into Petri dishes.

Appendix 2: Relevant DNA and Protein Sequences

2.1 Expression plasmids

- [pRSETA - 2897 bp] pRSETA complete sequence
  highlighted in green: *Bam*HI and *Eco*RI sites

```
   1 GATCTCGATC CCGCGAAATT AATACGACTC ACTATAGGGA GACCACAACG
  51 GTTTCCCTCT AGAAATAATT TTGTTTAACT TTAAGAAGGA GATATACATA
 101 TGCGGGGTTC TCATCATCAT CATCATCATG GTATGGCTAG CATGACTGGT
 151 GGACAGCAAA TGGGTCGGGA TCTGTACGAC GATGACGATA AGGATCGATG
 201 GGGATCCGAG CTCGAGATCT GCAGCTGGTA CCATGGAATT CGAAGCTTGA
 251 TCCGGCTGCT AACAAAGCCC GAAAGGAAGC TGAGTTGGCT GCTGCCACCG
 301 CTGAGCAATA ACTAGCATAA CCCCTTGGGG CCTCTAAACG GGTCTTGAGG
 351 GGTTTTTTGC TGAAAGGAGG AACTATATCC GGATCTGGCG TAATAGCGAA
 401 GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC TGAATGGCGA
 451 ATGGGACGCG CCCTGTAGCG GCGCATTAAG CGCGGCGGGT GTGGTGGTTA
 501 CGCGCAGCGT GACCGCTACA CTTGCCAGCG CCCTAGCGCC CGCTCCTTTC
 551 GCTTTCTTCC CTTCCTTTCT CGCCACGTTC GCCGGCTTTC CCCGTCAAGC
 601 TCTAAATCGG GGGCTCCCTT TAGGGTTCCG ATTTAGTGCT TTACGGCACC
 651 TCGACCCCAA AAAACTTGAT TAGGGTGATG GTTCACGTAG TGGGCCATCG
 701 CCCTGATAGA CGGTTTTTCG CCCTTTGACG TTGGAGTCCA CGTTCTTTAA
 751 TAGTGGACTC TTGTTCCAAA CTGGAACAAC ACTCAACCCT ATCTCGGTCT
 801 ATTCTTTTGA TTTATAAGGG ATTTTGCCGA TTTCGGCCTA TTGGTTAAAA
 851 AATGAGCTGA TTTAACAAAA ATTTAACGCG AATTTTAACA AAATATTAAC
 901 GCTTACAATT TAGGTGGCAC TTTTCGGGGA AATGTGCGCG GAACCCCTAT
 951 TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGAGACAAT
1001 AACCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT
1051 CAACATTTCC GTGTCGCCCT TATTCCCTTT TTTGCGGCAT TTTGCCTTCC
1101 TGTTTTTGCT CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC
1151 AGTTGGGTGC ACGAGTGGGT TACATCGAAC TGGATCTCAA CAGCGGTAAG
1201 ATCCTTGAGA GTTTTCGCCC CGAAGAACGT TTTCCAATGA TGAGCACTTT
1251 TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTATTGAC GCCGGGCAAG
1301 AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTTGAGTAC
1351 TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT
1401 ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC
1451 TGACAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTT GCACAACATG
```

```
1501 GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC TGAATGAAGC
1551 CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA
1601 CGTTGCGCAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA
1651 CAATTAATAG ACTGGATGGA GGCGGATAAA GTTGCAGGAC CACTTCTGCG
1701 CTCGGCCCTT CCGGCTGGCT GGTTTATTGC TGATAAATCT GGAGCCGGTG
1751 AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC
1801 TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA CTATGGATGA
1851 ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT AAGCATTGGT
1901 AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACTT
1951 CATTTTTAAT TTAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT
2001 GACCAAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG
2051 TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTTTTTCT GCGCGTAATC
2101 TGCTGCTTGC AAACAAAAAA ACCACCGCTA CCAGCGGTGG TTTGTTTGCC
2151 GGATCAAGAG CTACCAACTC TTTTTCCGAA GGTAACTGGC TTCAGCAGAG
2201 CGCAGATACC AAATACTGTT CTTCTAGTGT AGCCGTAGTT AGGCCACCAC
2251 TTCAAGAACT CTGTAGCACC GCCTACATAC CTCGCTCTGC TAATCCTGTT
2301 ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT
2351 CAAGACGATA GTTACCGGAT AAGGCGCAGC GGTCGGGCTG AACGGGGGGT
2401 TCGTGCACAC AGCCCAGCTT GGAGCGAACG ACCTACACCG AACTGAGATA
2451 CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG
2501 CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCACGAGG
2551 GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG
2601 CCACCTCTGA CTTGAGCGTC GATTTTTGTG ATGCTCGTCA GGGGGGCGGA
2651 GCCTATGGAA AAACGCCAGC AACGCGGCCT TTTTACGGTT CCTGGCCTTT
2701 TGCTGGCCTT TTGCTCACAT GTTCTTTCCT GCGTTATCCC CTGATTCTGT
2751 GGATAACCGT ATTACCGCCT TTGAGTGAGC TGATACCGCT CGCCGCAGCC
2801 GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA
2851 ATACGCAAAC CGCCTCTCCC CGCGCGTTGG CCGATTCATT AATGCAG
```

- [pFastBac1 - 4776 bp] pFastBac1 complete sequence
  highlighted in green: *Sac*I and *Hin*dIII sites

```
   1 GACGCGCCCT GTAGCGGCGC ATTAAGCGCG GCGGGTGTGG TGGTTACGCG
  51 CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT
 101 TCTTCCCTTC CTTTCTCGCC ACGTTCGCCG GCTTTCCCCG TCAAGCTCTA
 151 AATCGGGGGC TCCCTTTAGG GTTCCGATTT AGTGCTTTAC GGCACCTCGA
 201 CCCCAAAAAA CTTGATTAGG GTGATGGTTC ACGTAGTGGG CCATCGCCCT
 251 GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT CTTTAATAGT
 301 GGACTCTTGT TCCAAACTGG AACAACACTC AACCCTATCT CGGTCTATTC
 351 TTTTGATTTA TAAGGGATTT TGCCGATTTC GGCCTATTGG TTAAAAAATG
 401 AGCTGATTTA ACAAAAATTT AACGCGAATT TTAACAAAAT ATTAACGTTT
 451 ACAATTTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG
 501 TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC
 551 CCTGATAAAT GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA
 601 CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCCTGT
 651 TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
```

```
 701 TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC
 751 CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA
 801 AGTTCTGCTA TGTGGCGCGG TATTATCCCG TATTGACGCC GGGCAAGAGC
 851 AACTCGGTCG CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA
 901 CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG
 951 CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA
1001 CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG
1051 GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT
1101 ACCAAACGAC GAGCGTGACA CCACGATGCC TGTAGCAATG GCAACAACGT
1151 TGCGCAAACT ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA
1201 TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC
1251 GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC
1301 GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC
1351 CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG
1401 AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC
1451 TGTCAGACCA AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT
1501 TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA ATCTCATGAC
1551 CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG
1601 AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC
1651 TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA
1701 TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC
1751 AGATACCAAA TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC
1801 AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC
1851 AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA
1901 GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG
1951 TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT
2001 ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG
2051 ACAGGTATCC GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG
2101 CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG GGTTTCGCCA
2151 CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC
2201 TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC
2251 TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA
2301 TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA
2351 CGACCGAGCG CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG
2401 CGGTATTTTC TCCTTACGCA TCTGTGCGGT ATTTCACACC GCAGACCAGC
2451 CGCGTAACCT GGCAAAATCG GTTACGGTTG AGTAATAAAT GGATGCCCTG
2501 CGTAAGCGGG TGTGGGCGGA CAATAAAGTC TTAAACTGAA CAAAATAGAT
2551 CTAAACTATG ACAATAAAGT CTTAAACTAG ACAGAATAGT TGTAAACTGA
2601 AATCAGTCCA GTTATGCTGT GAAAAAGCAT ACTGGACTTT TGTTATGGCT
2651 AAAGCAAACT CTTCATTTTC TGAAGTGCAA ATTGCCCGTC GTATTAAAGA
2701 GGGGCGTGGC CAAGGGCATG GTAAAGACTA TATTCGCGGC GTTGTGACAA
2751 TTTACCGAAC AACTCCGCGG CCGGGAAGCC GATCTCGGCT TGAACGAATT
2801 GTTAGGTGGC GGTACTTGGG TCGATATCAA AGTGCATCAC TTCTTCCCGT
2851 ATGCCCAACT TTGTATAGAG AGCCACTGCG GGATCGTCAC CGTAATCTGC
2901 TTGCACGTAG ATCACATAAG CACCAAGCGC GTTGGCCTCA TGCTTGAGGA
```

2951 GATTGATGAG CGCGGTGGCA ATGCCCTGCC TCCGGTGCTC GCCGGAGACT
3001 GCGAGATCAT AGATATAGAT CTCACTACGC GGCTGCTCAA ACCTGGGCAG
3051 AACGTAAGCC GCGAGAGCGC CAACAACCGC TTCTTGGTCG AAGGCAGCAA
3101 GCGCGATGAA TGTCTTACTA CGGAGCAAGT TCCCGAGGTA ATCGGAGTCC
3151 GGCTGATGTT GGGAGTAGGT GGCTACGTCT CCGAACTCAC GACCGAAAAG
3201 ATCAAGAGCA GCCCGCATGG ATTTGACTTG GTCAGGGCCG AGCCTACATG
3251 TGCGAATGAT GCCCATACTT GAGCCACCTA ACTTTGTTTT AGGGCGACTG
3301 CCCTGCTGCG TAACATCGTT GCTGCTGCGT AACATCGTTG CTGCTCCATA
3351 ACATCAAACA TCGACCCACG GCGTAACGCG CTTGCTGCTT GGATGCCCGA
3401 GGCATAGACT GTACAAAAAA ACAGTCATAA CAAGCCATGA AAACCGCCAC
3451 TGCGCCGTTA CCACCGCTGC GTTCGGTCAA GGTTCTGGAC CAGTTGCGTG
3501 AGCGCATACG CTACTTGCAT TACAGTTTAC GAACCGAACA GGCTTATGTC
3551 AACTGGGTTC GTGCCTTCAT CCGTTTCCAC GGTGTGCGTC ACCCGGCAAC
3601 CTTGGGCAGC AGCGAAGTCG AGGCATTTCT GTCCTGGCTG GCGAACGAGC
3651 GCAAGGTTTC GGTCTCCACG CATCGTCAGG CATTGGCGGC CTTGCTGTTC
3701 TTCTACGGCA AGGTGCTGTG CACGGATCTG CCCTGGCTTC AGGAGATCGG
3751 AAGACCTCGG CCGTCGCGGC GCTTGCCGGT GGTGCTGACC CCGGATGAAG
3801 TGGTTCGCAT CCTCGGTTTT CTGGAAGGCG AGCATCGTTT GTTCGCCCAG
3851 GACTCTAGCT ATAGTTCTAG TGGTTGGCTA CGTATACTCC GGAATATTAA
3901 TAGATCATGG AGATAATTAA AATGATAACC ATCTCGCAAA TAAATAAGTA
3951 TTTTACTGTT TTCGTAACAG TTTTGTAATA AAAAAACCTA TAAATATTCC
4001 GGATTATTCA TACCGTCCCA CCATCGGGCG CGGATCCCGG TCCGAAGCGC
4051 GCGGAATTCA AAGGCCTACG TCGACGAGCT CACTAGTCGC GGCCGCTTTC
4101 GAATCTAGAG CCTGCAGTCT CGAGGCATGC GGTACCAAGC TTGTCGAGAA
4151 GTACTAGAGG ATCATAATCA GCCATACCAC ATTTGTAGAG GTTTTACTTG
4201 CTTTAAAAAA CCTCCCACAC CTCCCCCTGA ACCTGAAACA TAAAATGAAT
4251 GCAATTGTTG TTGTTAACTT GTTTATTGCA GCTTATAATG GTTACAAATA
4301 AAGCAATAGC ATCACAAATT TCACAAATAA AGCATTTTTT TCACTGCATT
4351 CTAGTTGTGG TTTGTCCAAA CTCATCAATG TATCTTATCA TGTCTGGATC
4401 TGATCACTGC TTGAGCCTAG GAGATCCGAA CCAGATAAGT GAAATCTAGT
4451 TCCAAACTAT TTTGTCATTT TTAATTTTCG TATTAGCTTA CGACGCTACA
4501 CCCAGTTCCC ATCTATTTTG TCACTCTTCC CTAAATAATC CTTAAAAACT
4551 CCATTTCCAC CCCTCCCAGT TCCCAACTAT TTTGTCCGCC CACAGCGGGG
4601 CATTTTTCTT CCTGTTATGT TTTTAATCAA ACATCCTGCC AACTCCATGT
4651 GACAAACCGT CATCTTCGGC TACTTTTTCT CTGTCACAGA ATGAAAATTT
4701 TTCTGTCATC TCTTCGTTAT TAATGTTTGT AATTGACTGA ATATCAACGC
4751 TTATTTGCAG CCTGAATGGC GAATGG

*Sac*I-*Hin*dIII gene inserts for recombinant pFastBac1

- His-tagged 4har gene

  4076 gagctcacaa cggtttccct ctagaaataa ttttgtttaa ctttaagaag
  4126 gagatataca tatgcggggt tctCATCATC ATCATCATCA Tggtatggct
  4176 agcatgactg gtggacagca aatgggtcgg gatctgtacg acgatgacga
  4226 taaggatcga tggGGATCCG ACCCAGCGGC ACTGAAACGT GCACGCAACA
  4276 CCGAAGCTGC ACGTCGTTCC CGTGCTCGTA AACTGCAGCG TATGAAACAA

```
4326 CTGGAAGACA AAGTTGAAGA GCTGCTGTCC AAGAACTACC ACCTGGAAAA
4376 CGAAGTTGCT CGTCTGAAAA AACTGGTTGC GCGCCTGGCG CGTCAAGTAC
4426 GTGCTCTGGC TGATTCTCTG ATGCAGCTGG CTCGCCAGGT TTCCCGTCTG
4476 GCAGACTCCC TGTAGTAGAA TTCGAAGCTT
```

– His-tagged 4hee gene

```
4076 GAGCTCACAA CGGTTTCCct ctagaaataa ttttgtttaa ctttaagaag
4126 gagatataca tatgcggggt tctCATCATC ATCATCATCA Tggtatggct
4176 agcatgactg gtggacagca aatgggtcgg gatctgtacg acgatgacga
4226 taaggatcga tggGGATCCG ACCCAGCGGC ACTGAAACGT GCACGCAACA
4276 CCGAAGCTGC ACGTCGTTCC CGTGCTCGTA AACTGCAGCG TATGAAACAA
4326 CTGGAAGACA AAGTTGAAGA GCTGCTGTCC AAGAACTACC ACCTGGAAAA
4376 CGAAGTTGCT CGTCTGAAAA AACTGGTTGA GGAGCTGGCG CGTCAAGTAC
4426 GTGCTCTGGC TGATTCTCTG ATGCAGCTGG CTCGCCAGGT TTCCCGTCTG
4476 GCAGACTCCC TGTAGTAGaa ttcgaAGCTT
```

– His-tagged lzar gene

```
4076 GAGCTCACAA CGGTTTCCct ctagaaataa ttttgtttaa ctttaagaag
4126 gagatataca tatgcggggt tctcatcatc atcatcatca tggtatggct
4176 agcatgactg gtggacagca aatgggtcgg gatctgtacg acgatgacga
4226 taaggatcga tggGGATCCG ACCCAGCGGC ACTGAAACGT GCACGCAACA
4276 CCGAAGCTGC ACGTCGTTCC CGTGCTCGTA AACTGCAGCG TATGAAACAA
4326 CTGGAAGACA AAGTTGAAGA GCTGCTGTCC AAGAACTACC ACCTGGAAAA
4376 CGAAGTTGCT CGTCTGAAAA AACTGGTTGC GCGCCTGAAG AAGCTGGTAC
4426 GTGCTCTGGC TGATTCTCTG ATGCAGCTGG CTCGCCAGGT TTCCCGTCTG
4476 GAATCCGGTC AGTAGTAGaa ttcgaagctt
```

– His-tagged lzee gene

```
4076 GAGCTCACAA CGGTTTCCct ctagaaataa ttttgtttaa ctttaagaag
4126 gagatataca tatgcggggt tctcatcatc atcatcatca tggtatggct
4176 agcatgactg gtggacagca aatgggtcgg gatctgtacg acgatgacga
4226 taaggatcga tggGGATCCG ACCCAGCGGC ACTGAAACGT GCACGCAACA
4276 CCGAAGCTGC ACGTCGTTCC CGTGCTCGTA AACTGCAGCG TATGAAACAA
4326 CTGGAAGACA AAGTTGAAGA GCTGCTGTCC AAGAACTACC ACCTGGAAAA
4376 CGAAGTTGCT CGTCTGAAAA AACTGGTTGA GGAGCTGCTG TCCAAAGTAC
4426 GTGCTCTGGC TGATTCTCTG ATGCAGCTGG CTCGCCAGGT TTCCCGTCTG
4476 GAATCCGGTC AGTAGTAGaa ttcgaAGCTT
```

- [pGEX6P1 - 4984 bp] pGEX6P1 complete sequence
  uppercase: P_{tac} (-35 and -10), GST affinity tag, multiple cloning site, *lacI*, $P_{wt\text{-}lac}$ (-35 and -10), lacZα
  highlighted in green: *Eco*NI, *Eco*RI and *Bam*HI sites
  *sequence in cyan was changed to lacUV5 and wild-type lac promoters
  *sequence in yellow was deleted in lacUV5-pGEX6P1b plasmids

```
   1 acgttatcga ctgcacggtg caccaatgct tctggcgtca ggcagccatc
  51 ggaagctgtg gtatggctgt gcaggtcgta aatcactgca taattcgtgt
 101 cgctcaaggc gcactcccgt tctggataat gtttttgcg ccgacatcat
```

```
 151 aacggttctg gcaaatattc tgaaatgagc tgTTGACAat taatcatcgg
 201 ctcgTATAAT gtgtggaatt gtgagcggat aacaatttca cacaggaaac
 251 agtattcATG TCCCCTATAC TAGGTTATTG GAAAATTAAG GGCCTTGTGC
 301 AACCCACTCG ACTTCTTTTG GAATATCTTG AAGAAAAATA TGAAGAGCAT
 351 TTGTATGAGC GCGATGAAGG TGATAAATGG CGAAACAAAA AGTTTGAATT
 401 GGGTTTGGAG TTTCCCAATC TTCCTTATTA TATTGATGGT GATGTTAAAT
 451 TAACACAGTC TATGGCCATC ATACGTTATA TAGCTGACAA GCACAACATG
 501 TTGGGTGGTT GTCCAAAAGA GCGTGCAGAG ATTTCAATGC TTGAAGGAGC
 551 GGTTTTGGAT ATTAGATACG GTGTTTCGAG AATTGCATAT AGTAAAGACT
 601 TTGAAACTCT CAAAGTTGAT TTTCTTAGCA AGCTACCTGA AATGCTGAAA
 651 ATGTTCGAAG ATCGTTTATG TCATAAAACA TATTTAAATG GTGATCATGT
 701 AACCCATCCT GACTTCATGT TGTATGACGC TCTTGATGTT GTTTTATACA
 751 TGGACCCAAT GTGCCTGGAT GCGTTCCCAA AATTAGTTTG TTTTAAAAAA
 801 CGTATTGAAG CTATCCCACA AATTGATAAG TACTTGAAAT CCAGCAAGTA
 851 TATAGCATGG CCTTTGCAGG GCTGGCAAGC CACGTTTGGT GGTGGCGACC
 901 ATCCTCCAAA ATCGGATctg gaagttctgt tccaggggcc cctgGGATCC
 951 CCGGAATTCC CGGGTCGACT CGAGCGGCCG Catcgtgact gactgacgat
1001 ctgcctcgcg cgtttcggtg atgacggtga aaacctctga cacatgcagc
1051 tcccggagac ggtcacagct tgtctgtaag cggatgccgg gagcagacaa
1101 gcccgtcagg gcgcgtcagc gggtgttggc gggtgtcggg gcgcagccat
1151 gacccagtca cgtagcgata gcggagtgta taattcttga agacgaaagg
1201 gcctcgtgat acgcctattt ttataggtta atgtcatgat aataatggtt
1251 tcttagacgt caggtggcac ttttcgggga aatgtgcgcg gaacccctat
1301 ttgtttattt ttctaaatac attcaaatat gtatccgctc atgagacaat
1351 aaccctgata aatgcttcaa taatattgaa aaaggaagag tatgagtatt
1401 caacatttcc gtgtcgccct tattcccttt tttgcggcat tttgccttcc
1451 tgtttttgct cacccagaaa cgctggtgaa agtaaaagat gctgaagatc
1501 agttgggtgc acgagtgggt tacatcgaac tggatctcaa cagcggtaag
1551 atccttgaga gttttcgccc cgaagaacgt tttccaatga tgagcacttt
1601 taaagttctg ctatgtggcg cggtattatc ccgtgttgac gccgggcaag
1651 agcaactcgg tcgccgcata cactattctc agaatgactt ggttgagtac
1701 tcaccagtca cagaaaagca tcttacggat ggcatgacag taagagaatt
1751 atgcagtgct gccataacca tgagtgataa cactgcggcc aacttacttc
1801 tgacaacgat cggaggaccg aaggagctaa ccgcttttt gcacaacatg
1851 ggggatcatg taactcgcct tgatcgttgg gaaccggagc tgaatgaagc
1901 cataccaaac gacgagcgtg acaccacgat gcctgcagca atggcaacaa
1951 cgttgcgcaa actattaact ggcgaactac ttactctagc ttcccggcaa
2001 caattaatag actggatgga ggcggataaa gttgcaggac cacttctgcg
2051 ctcggccctt ccggctggct ggtttattgc tgataaatct ggagccggtg
2101 agcgtgggtc tcgcggtatc attgcagcac tggggccaga tggtaagccc
2151 tcccgtatcg tagttatcta cacgacgggg agtcaggcaa ctatggatga
2201 acgaaataga cagatcgctg agataggtgc ctcactgatt aagcattggt
2251 aactgtcaga ccaagtttac tcatatatac tttagattga tttaaaactt
2301 catttttaat ttaaaaggat ctaggtgaag atccttttg ataatctcat
2351 gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg tcagaccccg
```

2401 tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc
2451 tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc
2501 ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag
2551 cgcagatacc aaatactgtc cttctagtgt agccgtagtt aggccaccac
2601 ttcaagaact ctgtagcacc gcctacatac ctcgctctgc taatcctgtt
2651 accagtggct gctgccagtg gcgataagtc gtgtcttacc gggttggact
2701 caagacgata gttaccggat aaggcgcagc ggtcgggctg aacgggggggt
2751 tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata
2801 cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg
2851 cggacaggta tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg
2901 gagcttccag ggggaaacgc ctggtatctt tatagtcctg tcgggtttcg
2951 ccacctctga cttgagcgtc gatttttgtg atgctcgtca gggggggcgga
3001 gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt
3051 tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt
3101 ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc
3151 gaacgaccga gcgcagcgag tcagtgagcg aggaagcgga agagcgcctg
3201 atgcggtatt ttctccttac gcatctgtgc ggtatttcac accgcataaa
3251 ttccgacacc atcgaatggt gcaaaacctt tcgcggtatg gcatgatagc
3301 gcccggaaga gagtcaattc agggtggtga atGTGAAACC AGTAACGTTA
3351 TACGATGTCG CAGAGTATGC CGGTGTCTCT TATCAGACCG TTTCCCGCGT
3401 GGTGAACCAG GCCAGCCACG TTTCTGCGAA AACGCGGGAA AAAGTGGAAG
3451 CGGCGATGGC GGAGCTGAAT TACATTCCCA ACCGCGTGGC ACAACAACTG
3501 GCGGGCAAAC AGTCGTTGCT GATTGGCGTT GCCACCTCCA GTCTGGCCCT
3551 GCACGCGCCG TCGCAAATTG TCGCGGCGAT TAAATCTCGC GCCGATCAAC
3601 TGGGTGCCAG CGTGGTGGTG TCGATGGTAG AACGAAGCGG CGTCGAAGCC
3651 TGTAAAGCGG CGGTGCACAA TCTTCTCGCG CAACGCGTCA GTGGGCTGAT
3701 CATTAACTAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCT
3751 GCACTAATGT TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC
3801 AACAGTATTA TTTTCTCCCA TGAAGACGGT ACGCGACTGG GCGTGGAGCA
3851 TCTGGTCGCA TTGGGTCACC AGCAAATCGC GCTGTTAGCG GGCCCATTAA
3901 GTTCTGTCTC GGCGCGTCTG CGTCTGGCTG GCTGGCATAA ATATCTCACT
3951 CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT GGAGTGCCAT
4001 GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA
4051 CTGCGATGCT GGTTGCCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC
4101 ATTACCGAGT CCGGGCTGCG CGTTGGTGCG GATATCTCGG TAGTGGGATA
4151 CGACGATACC GAAGACAGCT CATGTTATAT CCCGCCGTCA ACCACCATCA
4201 AACAGGATTT TCGCCTGCTG GGGCAAACCA GCGTGGACCG CTTGCTGCAA
4251 CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC CCGTCTCACT
4301 GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC
4351 GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG
4401 GAAAGCGGGC AGTGAgcgca acgcaattaa tgtgagttag ctcactcatt
4451 aggcacccca ggcTTTACAc tttatgcttc cggctcgTAT GTTgtgtgga
4501 attgtgagcg gataacaatt tcacacagga aacagctATG ACCATGATTA
4551 CGGATTCACT GGCCGTCGTT TTACAACGTC GTGACTGGGA AAACCCTGGC
4601 GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG

```
4651 TAATAGCGAA GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC
4701 TGAATGGCGA ATGGCGCTTT GCCTGGTTTC CGGCACCAGA AGCGGTGCCG
4751 GAAAGCTGGC TGGAGTGCGA TCTTCCTGAG GCCGATACTG TCGTCGTCCC
4801 CTCAAACTGG CAGATGCACG GTTACGATGC GCCCATCTAC ACCAACGTAA
4851 CCTATCCCAT TACGGTCAAT CCGCCGTTTG TTCCCACGGA GAATCCGACG
4901 GGTTGTTACT CGCTCACATT TAATGTTGAT GAAAGCTGGC TACAGGAAGG
4951 CCAGACGCGA ATTATTTTTG ATGGCGTTgg TAATGTTGAT GAAAGCTGGC
5001 TACAGGAAGG CCAGACGCGA ATTATTTTTG ATGGCGTTgg aatt
```

Modifications on the promoter sequence
uppercase sequences: -35 and -10

- [lacUV5-pGEX6P1 - 4986 bp] lacUV5 promoter sequence
  ```
  183 TTTACActtt atgcttccgg ctcgTATAAT
  ```
- [lacWT-pGEX6P1 - 4986 bp] wild-type lac promoter sequence
  ```
  183 TTTACActtt atgcttccgg ctcgTATGTT
  ```

*Bam*HI-*Eco*RI gene inserts for recombinant pRSETA, pGEX6P1 and pGEX6P1b plasmids

- 4har gene (in pRSETA and pGEX6P1)
  ```
    1 GGATCCGACC CAGCGGCACT GAAACGTGCA CGCAACACCG AAGCTGCACG
   51 TCGTTCCCGT GCTCGTAAAC TGCAGCGTAT GAAACAACTG GAAGACAAAG
  101 TTGAAGAGCT GCTGTCCAAG AACTACCACC TGGAAAACGA AGTTGCTCGT
  151 CTGAAAAAAC TGGTTGCGCG CCTGGCGCGT CAAGTACGTG CTCTGGCTGA
  201 TTCTCTGATG CAGCTGGCTC GCCAGGTTTC CCGTCTGGCA GACTCCCTGT
  251 AGTAGAATTC
  ```
- 4hee gene (in pRSETA)
  ```
    1 GGATCCGACC CAGCGGCACT GAAACGTGCA CGCAACACCG AAGCTGCACG
   51 TCGTTCCCGT GCTCGTAAAC TGCAGCGTAT GAAACAACTG GAAGACAAAG
  101 TTGAAGAGCT GCTGTCCAAG AACTACCACC TGGAAAACGA AGTTGCTCGT
  151 CTGAAAAAAC TGGTTGAGGA GCTGGCGCGT CAAGTACGTG CTCTGGCTGA
  201 TTCTCTGATG CAGCTGGCTC GCCAGGTTTC CCGTCTGGCA GACTCCCTGT
  251 AGTAGAATTC
  ```
- lzar gene (in pRSETA)
  ```
    1 GGATCCGACC CAGCGGCACT GAAACGTGCA CGCAACACCG AAGCTGCACG
   51 TCGTTCCCGT GCTCGTAAAC TGCAGCGTAT GAAACAACTG GAAGACAAAG
  101 TTGAAGAGCT GCTGTCCAAG AACTACCACC TGGAAAACGA AGTTGCTCGT
  151 CTGAAAAAAC TGGTTGCGCG CCTGAAGAAG CTGGTACGTG CTCTGGCTGA
  201 TTCTCTGATG CAGCTGGCTC GCCAGGTTTC CCGTCTGGAA TCCGGTCAGT
  251 AGTAGAATTC
  ```
- lzee gene (in pRSETA)
  highlighted in yellow: base in original sequence design
  ```
  202 GGATCCGACC CAGCGGCACT GAAACGTGCA CGCAACACCG AAGCTGCACG
  252 TCGTTCCCGT GCTCGTAAAC TGCAGCGTAT GAAACAACTG GAAGACAAAG
  ```

```
 302 TTGAAGAGCT GCTGTCCAAG AACTACCACC TGGAAAACGA AGTTGCTCGT
 352 CTGAAAAAAC TGGTTGAGGA GCTGCTGTCC AAAGTACGTG CTCTGGCTGA
 402 TTCTCTGATG CAGCTGGCTC GCCAGGTTTC CCGTCTGGAA TCCGGTCAGT
 452 AGTAGAATTC
```

– lzee gene (in pGEX6P1)
highlighted in yellow: point mutation resulting to Ser instead of Arg

```
 945 GGATCCGACC CAGCGGCACT GAAACGTGCA CGCAACACCG AAGCTGCACG
 995 TCGTTCCCGT GCTCGTAAAC TGCAGCGTAT GAAACAACTG GAAGACAAAG
1045 TTGAAGAGCT GCTGTCCAAG AACTACCACC TGGAAAACGA AGTTGCTCGT
1095 CTGAAAAAAC TGGTTGAGGA GCTGCTGTCC AAAGTAAGTG CTCTGGCTGA
1145 TTCTCTGATG CAGCTGGCTC GCCAGGTTTC CCGTCTGGAA TCCGGTCAGT
1195 AGTAGAATTC
```

– [lacUV5-lzee.v2/pGEX6P1 - 5231 bp] lzee gene with added restriction enzyme sites
highlighted in green: *Spe*I and *Mfe*I sites
highlighted in yellow: base in original sequence design

```
 947 GGATCCGACC CAGCGGCACT GAAACGTGCA CGCAACACCG AAGCTGCACG
 997 TCGTTCCCGT GCTCGTAAAC TGCAGCGTAT GAAACAACTG GAAGACAAAG
1047 TTGAAGAGCT GCTGTCCAAG AACTACCACC TGGAAAACGA AGTTGCTCGT
1097 CTGAAAAAAC TAGTTGAGGA GCTGCTGTCC AAAGTACGTG CTCTGGCTGA
1147 TTCTCTGATG CAATTGGCTC GCCAGGTTTC CCGTCTGGAA TCCGGTCAGT
1197 AGTAGAATTC
```

– [lacUV5-lzee.v2 libctrl/pGEX6P1 - 5231 bp] lzee.v2 library control gene
highlighted in green: *Spe*I, *Sac*I, *Hin*dIII and *Mfe*I sites

```
 947 GGATCCGACC CAGCGGCACT GAAACGTGCA CGCAACACCG AAGCTGCACG
 997 TCGTTCCCGT GCTCGTAAAC TGCAGCGTAT GAAACAACTG GAAGACAAAG
1047 TTGAAGAGCT GCTGTCCAAG AACTACCACC TGGAAAACGA AGTTGCTCGT
1097 CTGAAAAAAC TAGTTGAGGA GCTCCTGTCC AAGCTTCGTG CTCTGGCTGA
1147 TTCTCTGATG CAATTGGCTC GCCAGGTTTC CCGTCTGGAA TCCGGTCAGT
1197 AGTAGAATTC
```

– LZD73 gene

```
   1 GGATCCGATC CAGCTGCTCT GAAGCGAGCT CGGAACACTG AAGCTGCTCG
  51 ACGGAGCCGA GCTCGGAAGC TGCAACGAAT GAAGCAGCTG GAAGACAAGG
 101 TGGAGGAACT GCTGAGCAAG AACTACCACC TGGAGAACGA AGTTGCGCGC
 151 CTGAAGAAGC TGGTGGGTGA ACTGCAGAAG TTACAGCGGG TGAAGCGAGC
 201 TCGGAACACT GAAGCTGCTC GACGGAGCCG AGCTCGAAAG GCTGCTCTGA
 251 AGGGATAGTA AGAATTC
```

– LZD80 gene

```
   1 GGATCCGATC CAGCTGCTCT GAAGCGAGCT CGGAACACTG AAGCTGCTCG
  51 ACGGAGCCGA GCTCGGAAGC TGCAACGAAT GAAGCAGCTG GAAGACAAGG
 101 TGGAGGAACT GCTGAGCAAG AACTACCACC TGGAGAACGA AGTTGCGCGC
 151 CTGAAAAAGC TGGTGGAAGA ACTGCTGAGC AAAGTGGGCG AACTGCAGAA
 201 GTTACAGCGG GTGAAGCGAG CTCGGAACAC TGAAGCTGCT CGACGGAGCC
```

```
251 GAGCTCGAAA GGCTGCTCTG AAGGGATAGT AAGAATTC
```

– LZD87 gene

```
  1 GGATCCGATC CAGCTGCTCT GAAGCGAGCT CGGAACACTG AAGCTGCTCG
 51 ACGGAGCCGA GCTCGGAAGC TGCAACGAAT GAAGCAGCTG GAAGACAAGG
101 TGGAGGAACT GCTGAGCAAG AACTACCACC TGGAGAACGA AGTTGCGCGC
151 CTGAAAAAGC TGGTGGAAGA ACTGCTGAGC AAAGTGCGTG CGCTGGCGGA
201 TTCTCTGGGC GAACTGCAGA AGTTACAGCG GGTGAAGCGA GCTCGGAACA
251 CTGAAGCTGC TCGACGGAGC CGAGCTCGAA AGGCTGCTCT GAAGGGATAG
301 TAAGAATTC
```

- [pBAD/Myc-HisA - 4094 bp] pBAD/Myc-HisA complete sequence highlighted in green: *Kpn*I and *Eco*RI sites

```
   1 AAGAAACCAA TTGTCCATAT TGCATCAGAC ATTGCCGTCA CTGCGTCTTT
  51 TACTGGCTCT TCTCGCTAAC CAAACCGGTA ACCCCGCTTA TTAAAAGCAT
 101 TCTGTAACAA AGCGGGACCA AGCCATGAC AAAAACGCGT AACAAAAGTG
 151 TCTATAATCA CGGCAGAAAA GTCCACATTG ATTATTTGCA CGGCGTCACA
 201 CTTTGCTATG CCATAGCATT TTTATCCATA AGATTAGCGG ATCCTACCTG
 251 ACGCTTTTTA TCGCAACTCT CTACTGTTTC TCCATACCCG TTTTTTGGGC
 301 TAACAGGAGG AATTAACCAT GGATCCGAGC TCGAGATCTG CAGCTggtac
 351 catatgggaa ttcGAAGCTT GGGCCCGAAC AAAAACTCAT CTCAGAAGAG
 401 GATCTGAATA GCGCCGTCGA CCATCATCAT CATCATCATT GAGTTTAAAC
 451 GGTCTCCAGC TTGGCTGTTT TGGCGGATGA GAGAAGATTT TCAGCCTGAT
 501 ACAGATTAAA TCAGAACGCA GAAGCGGTCT GATAAAACAG AATTTGCCTG
 551 GCGGCAGTAG CGCGGTGGTC CCACCTGACC CCATGCCGAA CTCAGAAGTG
 601 AAACGCCGTA GCGCCGATGG TAGTGTGGGG TCTCCCCATG CGAGAGTAGG
 651 GAACTGCCAG GCATCAAATA AAACGAAAGG CTCAGTCGAA AGACTGGGCC
 701 TTTCGTTTTA TCTGTTGTTT GTCGGTGAAC GCTCTCCTGA GTAGGACAAA
 751 TCCGCCGGGA GCGGATTTGA ACGTTGCGAA GCAACGGCCC GGAGGGTGGC
 801 GGGCAGGACG CCCGCCATAA ACTGCCAGGC ATCAAATTAA GCAGAAGGCC
 851 ATCCTGACGG ATGGCCTTTT TGCGTTTCTA CAAACTCTTT TGTTTATTTT
 901 TCTAAATACA TTCAAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA
 951 ATGCTTCAAT AATATTGAAA AAGGAAGAGT ATGAGTATTC AACATTTCCG
1001 TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT GTTTTTGCTC
1051 ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA
1101 CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG
1151 TTTTCGCCCC GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC
1201 TATGTGGCGC GGTATTATCC CGTGTTGACG CCGGGCAAGA GCAACTCGGT
1251 CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC
1301 AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA TGCAGTGCTG
1351 CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC
1401 GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT
1451 AACTCGCCTT GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG
1501 ACGAGCGTGA CACCACGATG CCTGTAGCAA TGGCAACAAC GTTGCGCAAA
1551 CTATTAACTG GCGAACTACT TACTCTAGCT TCCCGGCAAC AATTAATAGA
1601 CTGGATGGAG GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC
```

1651 CGGCTGGCTG GTTTATTGCT GATAAATCTG GAGCCGGTGA GCGTGGGTCT
1701 CGCGGTATCA TTGCAGCACT GGGGCCAGAT GGTAAGCCCT CCCGTATCGT
1751 AGTTATCTAC ACGACGGGGA GTCAGGCAAC TATGGATGAA CGAAATAGAC
1801 AGATCGCTGA GATAGGTGCC TCACTGATTA AGCATTGGTA ACTGTCAGAC
1851 CAAGTTTACT CATATATACT TTAGATTGAT TTAAAACTTC ATTTTTAATT
1901 TAAAAGGATC TAGGTGAAGA TCCTTTTTGA TAATCTCATG ACCAAAATCC
1951 CTTAACGTGA GTTTTCGTTC CACTGAGCGT CAGACCCCGT AGAAAAGATC
2001 AAAGGATCTT CTTGAGATCC TTTTTTTCTG CGCGTAATCT GCTGCTTGCA
2051 AACAAAAAAA CCACCGCTAC CAGCGGTGGT TTGTTTGCCG GATCAAGAGC
2101 TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC GCAGATACCA
2151 AATACTGTCC TTCTAGTGTA GCCGTAGTTA GGCCACCACT TCAAGAACTC
2201 TGTAGCACCG CCTACATACC TCGCTCTGCT AATCCTGTTA CCAGTGGCTG
2251 CTGCCAGTGG CGATAAGTCG TGTCTTACCG GGTTGGACTC AAGACGATAG
2301 TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT CGTGCACACA
2351 GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG
2401 AGCTATGAGA AAGCGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT
2451 CCGGTAAGCG GCAGGGTCGG AACAGGAGAG CGCACGAGGG AGCTTCCAGG
2501 GGGAAACGCC TGGTATCTTT ATAGTCCTGT CGGGTTTCGC CACCTCTGAC
2551 TTGAGCGTCG ATTTTTGTGA TGCTCGTCAG GGGGGCGGAG CCTATGGAAA
2601 AACGCCAGCA ACGCGGCCTT TTTACGGTTC CTGGCCTTTT GCTGGCCTTT
2651 TGCTCACATG TTCTTTCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA
2701 TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGCAGCCG AACGACCGAG
2751 CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA GAGCGCCTGA TGCGGTATTT
2801 TCTCCTTACG CATCTGTGCG GTATTTCACA CCGCATATGG TGCACTCTCA
2851 GTACAATCTG CTCTGATGCC GCATAGTTAA GCCAGTATAC ACTCCGCTAT
2901 CGCTACGTGA CTGGGTCATG GCTGCGCCCC GACACCCGCC AACACCCGCT
2951 GACGCGCCCT GACGGGCTTG TCTGCTCCCG GCATCCGCTT ACAGACAAGC
3001 TGTGACCGTC TCCGGGAGCT GCATGTGTCA GAGGTTTTCA CCGTCATCAC
3051 CGAAACGCGC GAGGCAGCAG ATCAATTCGC GCGCGAAGGC GAAGCGGCAT
3101 GCATAATGTG CCTGTCAAAT GGACGAAGCA GGGATTCTGC AAACCCTATG
3151 CTACTCCGTC AAGCCGTCAA TTGTCTGATT CGTTACCAAT TATGACAACT
3201 TGACGGCTAC ATCATTCACT TTTTCTTCAC AACCGGCACG GAACTCGCTC
3251 GGGCTGGCCC CGGTGCATTT TTTAAATACC CGCGAGAAAT AGAGTTGATC
3301 GTCAAAACCA ACATTGCGAC CGACGGTGGC GATAGGCATC CGGGTGGTGC
3351 TCAAAAGCAG CTTCGCCTGG CTGATACGTT GGTCCTCGCG CCAGCTTAAG
3401 ACGCTAATCC CTAACTGCTG GCGGAAAAGA TGTGACAGAC GCGACGGCGA
3451 CAAGCAAACA TGCTGTGCGA CGCTGGCGAT ATCAAAATTG CTGTCTGCCA
3501 GGTGATCGCT GATGTACTGA CAAGCCTCGC GTACCCGATT ATCCATCGGT
3551 GGATGGAGCG ACTCGTTAAT CGCTTCCATG CGCCGCAGTA ACAATTGCTC
3601 AAGCAGATTT ATCGCCAGCA GCTCCGAATA GCGCCCTTCC CCTTGCCCGG
3651 CGTTAATGAT TTGCCCAAAC AGGTCGCTGA AATGCGGCTG GTGCGCTTCA
3701 TCCGGGCGAA AGAACCCCGT ATTGGCAAAT ATTGACGGCC AGTTAAGCCA
3751 TTCATGCCAG TAGGCGCGCG GACGAAAGTA AACCCACTGG TGATACCATT
3801 CGCGAGCCTC CGGATGACGA CCGTAGTGAT GAATCTCTCC TGGCGGGAAC
3851 AGCAAAATAT CACCCGGTCG GCAAACAAAT TCTCGTCCCT GATTTTTCAC

```
3901 CACCCCCTGA CCGCGAATGG TGAGATTGAG AATATAACCT TTCATTCCCA
3951 GCGGTCGGTC GATAAAAAAA TCGAGATAAC CGTTGGCCTC AATCGGCGTT
4001 AAACCCGCCA CCAGATGGGC ATTAAACGAG TATCCCGGCA GCAGGGGATC
4051 ATTTTGCGCT TCAGCCATAC TTTTCATACT CCCGCCATTC AGAG
```

*Kpn*I-*Eco*RI gene inserts for recombinant pBAD/Myc-HisA plasmids

- untagged lzee gene

```
346 GGTACCGAAG TTCTGTTCCA GGGGCCCCTG GGATCCGACC CAGCGGCACT
396 GAAACGTGCA CGCAACACCG AAGCTGCACG TCGTTCCCGT GCTCGTAAAC
446 TGCAGCGTAT GAAACAACTG GAAGACAAAG TTGAAGAGCT GCTGTCCAAG
496 AACTACCACC TGGAAAACGA AGTTGCTCGT CTGAAAAAAC TGGTTGAGGA
546 GCTGCTGTCC AAAGTACGTG CTCTGGCTGA TTCTCTGATG CAGCTGGCTC
596 GCCAGGTTTC CCGTCTGGAA TCCGGTCAGT AGTAGAATTC
```

- untagged LZD73 gene

```
346 GGTACCGAAG TTCTGTTCCA GGGGCCCCTG GGATCCGATC CAGCTGCTCT
396 GAAGCGAGCT CGGAACACTG AAGCTGCTCG ACGGAGCCGA GCTCGGAAGC
446 TGCAACGAAT GAAGCAGCTG GAAGACAAGG TGGAGGAACT GCTGAGCAAG
496 AACTACCACC TGGAGAACGA AGTTGCGCGC CTGAAGAAGC TGGTGGGTGA
546 ACTGCAGAAG TTACAGCGGG TGAAGCGAGC TCGGAACACT GAAGCTGCTC
596 GACGGAGCCG AGCTCGAAAG GCTGCTCTGA AGGGATAGTA AGAATTC
```

- untagged LZD80 gene

```
346 GGTACCGAAG TTCTGTTCCA GGGGCCCCTG GGATCCGATC CAGCTGCTCT
396 GAAGCGAGCT CGGAACACTG AAGCTGCTCG ACGGAGCCGA GCTCGGAAGC
446 TGCAACGAAT GAAGCAGCTG GAAGACAAGG TGGAGGAACT GCTGAGCAAG
496 AACTACCACC TGGAGAACGA AGTTGCGCGC CTGAAAAAGC TGGTGGAAGA
546 ACTGCTGAGC AAAGTGGGCG AACTGCAGAA GTTACAGCGG GTGAAGCGAG
596 CTCGGAACAC TGAAGCTGCT CGACGGAGCC GAGCTCGAAA GGCTGCTCTG
646 AAGGGATAGT AAGAATTC
```

- untagged LZD87 gene

```
346 GGTACCGAAG TTCTGTTCCA GGGGCCCCTG GGATCCGATC CAGCTGCTCT
396 GAAGCGAGCT CGGAACACTG AAGCTGCTCG ACGGAGCCGA GCTCGGAAGC
446 TGCAACGAAT GAAGCAGCTG GAAGACAAGG TGGAGGAACT GCTGAGCAAG
496 AACTACCACC TGGAGAACGA AGTTGCGCGC CTGAAAAAGC TGGTGGAAGA
546 ACTGCTGAGC AAAGTGCGTG CGCTGGCGGA TTCTCTGGGC GAACTGCAGA
596 AGTTACAGCG GGTGAAGCGA GCTCGGAACA CTGAAGCTGC TCGACGGAGC
646 CGAGCTCGAA AGGCTGCTCT GAAGGGATAG TAAGAATTC
```

2.2 Mutagenesis primer sequences for generating 300- to 1150-bp looping constructs with distal INV2 site

- [lambdaFP-EcoRI]

```
1 ccattcagct GAATTCcaga cgtaacagca ccac
```

- [lambdaRP-SacI]
```
   1 tttctgttcg GAGCTCccat taccacctta accg
```

- [mutFP-INV2-300]
```
   1 actagtATgt catatgacGA tgtacaCTGC GGGAAGTGCG GTATCAGCAT
  51 CGCAGAGC
```

- [mutRP-INV2-300]
```
   1 GACGCACTCC CCGCCGCCGC TGTTTTTGCG
```

- [mutFP-INV2-450]
```
   1 GTGCactagt TGgtcatatg acGGtgtaca CGGCAGAAAA TTCTGCCAGG
  51 GCGG
```

- [mutRP-INV2-450]
```
   1 TTGATGATGC GTTCGTTTCT GATGATTTTG CTGCCTCTTT TGAGGCCACC
  51 GCATCTCGTG
```

- [mutFP-INV2-650]
```
   1 actagtTCgt catatgacGC tgtacaCAGT GCCGGTGCGG CGAAAACGTC
  51 AG
```

- [mutRP-INV2-650]
```
   1 TGCGGCTGCG GCACTTTTTT CCGCTTCAGT GGCCTTTG
```

- [mutFP-INV2-900]
```
   1 actagtGGgt catatgacGG tgtacaACGC GTCCGTGGTG GCACAGAGTA
  51 CG
```

- [mutRP-INV2-900]
```
   1 TCAAGACGAC GCAGCACCTC CGGCCGG
```

- [mutFP3-BsrGI]
```
   1 tgtacaGTGG TGGTGAACAC GGTGGGCTCA G
```

- [mutRP-INV2-1150]
```
   1 GCgtcatatg acGgaattcT TGAAGACGAA AGGGCCTCGT
```

2.3 Promoter-operator constructs

- [pJ1653 - 5618 bp] pJ1653 complete sequence
  highlighted in green: *Eco*RI, *Sac*I, *Bam*HI, *Sal*I and *Not*I sites
  highlighted in cyan: -35 and -10 regions of lacUV5 promoter
```
   1 GAACTCCGGA TGAGCATTCA TCAGGCGGGC AAGAATGTGA ATAAAGGCCG
  51 GATAAAACTT GTGCTTATTT TTCTTTACGG TCTTTAAAAA GGCCGTAATA
 101 TCCAGCTGAA CGGTCTGGTT ATAGGTACAT TGAGCAACTG ACTGAAATGC
 151 CTCAAAATGT TCTTTACGAT GCCATTGGGA TATATCAACG GTGGTATATC
 201 CAGTGATTTT TTTCTCCATT TTAGCTTCCT TAGCTCCTGA AAATCTCGAT
```

```
 251 AACTCAAAAA ATACGCCCGG TAGTGATCTT ATTTCATTAT GGTGAAAGTT
 301 GGAACCTCTT ACGTGCCGAT CAACGTCTCA TTTTCGCCAA AAGTTGGCCC
 351 AGGGCTTCCC GGTATCAACA GGGACACCAG GATTTATTTA TTCTGCGAAG
 401 TGATCTTCCG TCACAGGTAT TTATTCGGCG CAAAGTGCGT CGGGTGATGC
 451 TGCCAACTTA CTGATTTAGT GTATGATGGT GTTTTTGAGG TGCTCCAGTG
 501 GCTTCTGTTT CTATCAGCTG TCCCTCCTGT TCAGCTACTG ACGGGGTGGT
 551 GCGTAACGGC AAAAGCACCG CCGGACATCA GCGCTAGCGG AGTGTATACT
 601 GGCTTACTAT GTTGGCACTG ATGAGGGTGT CAGTGAAGTG CTTCATGTGG
 651 CAGGAGAAAA AAGGCTGCAC CGGTGCGTCA GCAGAATATG TGATACAGGA
 701 TATATTCCGC TTCCTCGCTC ACTGACTCGC TACGCTCGGT CGTTCGACTG
 751 CGGCGAGCGG AAATGGCTTA CGAACGGGGC GGAGATTTCC TGGAAGATGC
 801 CAGGAAGATA CTTAACAGGG AAGTGAGAGG GCCGCGGCAA AGCCGTTTTT
 851 CCATAGGCTC CGCCCCCCTG ACAAGCATCA CGAAATCTGA CGCTCAAATC
 901 AGTGGTGGCG AAACCCGACA GGACTATAAA GATACCAGGC GTTTCCCCTG
 951 GCGGCTCCCT CGTGCGCTCT CCTGTTCCTG CCTTTCGGTT TACCGGTGTC
1001 ATTCCGCTGT TATGGCCGCG TTTGTCTCAT TCCACGCCTG ACACTCAGTT
1051 CCGGGTAGGC AGTTCGCTCC AAGCTGGACT GTATGCACGA ACCCCCCGTT
1101 CAGTCCGACC GCTGCGCCTT ATCCGGTAAC TATCGTCTTG AGTCCAACCC
1151 GGAAAGACAT GCAAAAGCAC CACTGGCAGC AGCCACTGGT AATTGATTTA
1201 GAGGAGTTAG TCTTGAAGTC ATGCGCCGGT TAAGGCTAAA CTGAAAGGAC
1251 AAGTTTTGGT GACTGCGCTC CTCCAAGCCA GTTACCTCGG TTCAAAGAGT
1301 TGGTAGCTCA GAGAACCTTC GAAAAACCGC CCTGCAAGGC GGTTTTTTCG
1351 TTTTCAGAGC AAGAGATTAC GCGCAGACCA AAACGATCTC AAGAAGATCA
1401 TCTTATTAAT CAGATAAAAT ATTTCTAGAT TTCAGTGCAA TTTATCTCTT
1451 CAAATGTAGC ACCTGAAGTC AGCCCCATAC GATATAAGTT GTAATTCTCA
1501 TGTTTGACAG CTTATCATCG ATAAGCTAGC TTTAATGCGG TAGTTTATCA
1551 CAGTTAAATT GCTAACGCAG TCAGGCACCG TGTATGAAAT CTAACAATGC
1601 GCTCATCGTC ATCCTCGGCA CCGTCACCCT GGATGCTGTA GGCATAGGCT
1651 TGGTTATGCC GGTTGCACAA TCTTCTCGCG CAACGCGTCA GTGGGCTGAT
1701 CATTAACTAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCT
1751 GCACTAATGT TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC
1801 AACAGTATTA TTTTCTCCCA TGAAGACGGT ACGCGACTGG GCGTGGAGCA
1851 TCTGGTCGCA TTGGGTCACC AGCAAATCGC GCTGTTAGCG GGCCCATTAA
1901 GTTCTGTCTC GGCGCGTCTG CGTCTGGCTG GCTGGCATAA ATATCTCACT
1951 CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT GGAGTGCCAT
2001 GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA
2051 CTGCGATGCT GGTTGCCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC
2101 ATTACCGAGT CCGGGCTGCG CGTTGGTGCG GATATCTCGG TAGTGGGATA
2151 CGACGATACC GAAGACAGCT CATGTTATAT CCCGCCGTTA ACCACCATCA
2201 AACAGGATTT TCGCCTGCTG GGGCAAACCA GCGTGGACCG CTTGCTGCAA
2251 CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC CCGTCTCACT
2301 GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC
2351 GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG
2401 GAAAGCGGGC AGTGAGCGCA ACGCAATTAA TGTAAGTTAG CGCGAATTAT
2451 CGTCCATTCC GACAGCATCG CCAGTCACTA TGGCGTGCTG CTAGCGCTAT
```

2501 ATGCGTTGAT GCAATTTCTA TGCGCACCCG TTCTCGGAGC ACTGTCCGAC
2551 CGCTTTGGCC GCCGCCCAGT CCTGCTCGCT TCGCTACTTG GAGCCACTAT
2601 CGACTACGCG ATCATGGCGA CCACACCCGT CCTGTGGGAT CAGATCTCGC
2651 AATTGCCGTT GATTTGGGGA TCTTTGTTGT AGGTGGACCA GTTGGTGATT
2701 TTGAACTTTT GCTTTGCCAC GGAACGGTCT GCGTTGTCGG GAAGATGCGT
2751 GATCTGATCC TTCAACTCAG CAAAAGTTCG ATTTATTCAA CAAAGCCGCC
2801 GTCCCGTCAA GTCAGCGTAA TGCTCTGCCA GTGTTACAAC CAATTAACCA
2851 ATTCTGATTA GAAAAACTCA TCGAGCATCA AATGAAACTG CAATTTATTC
2901 ATATCAGGAT TATCAATACC ATATTTTTGA AAAAGCCGTT TCTGTAATGA
2951 AGGAGAAAAC TCACCGAGGC AGTTCCATAG GATGGCAAGA TCCTGGTATC
3001 GGTCTGCGAT TCCGACTCGT CCAACATCAA TACAACCTAT TAATTTCCCC
3051 TCGTCAAAAA TAAGGTTATC AAGTGAGAAA TCACCATGAG TGACGACTGA
3101 ATCCGGTGAG AATGGCAAGA GCTTATGCAT TTCTTTCCAG ACTTGTTCAA
3151 CAGGCCAGCC ATTACGCTCG TCATCAAAAT CACTCGCATC AACCAAACCG
3201 TTATTCATTC GTGATTGCGC CTGAGCGAGA CGAAATACGC GATCGCTGTT
3251 AAAAGGACAA TTACAAACAG GAATCGAATG CAACCGGCGC AGGAACACTG
3301 CCAGCGCATC AACAATATTT TCACCTGAAT CAGGATATTC TTCTAATACC
3351 TGGAATGCTG TTTTCCCGGG GATCGCAGTG GTGAGTAACC ATGCATCATC
3401 AGGAGTACGG ATAAAATGCT TGATGGTCGG AAGAGGCATA AATTCCGTCA
3451 GCCAGTTTAG TCTGACCATC TCATCTGTAA CATCATTGGC AACGCTACCT
3501 TTGCCATGTT TCAGAAACAA CTCTGGCGCA TCGGGCTTCC CATACAATCG
3551 ATAGATTGTC GCACCTGATT GCCCGACATT ATCGCGAGCC CATTTATACC
3601 CATATAAATC AGCATCCATG TTGGAATTTA ATCGCGGCCT CGAGCAAGAC
3651 GTTTCCCGTT GAATATGGCT CATAACACCC CTTGTATTAC TGTTTATGTA
3701 AGCAGACAGT TTTATTGTTC ATGATGATAT ATTTTTATCT TGTGCAATGT
3751 AACATCAGAG ATTTTGAGAC ACAACGTGGC TTTGTTGAAT AAATCGAACT
3801 TTTGCTGAGT TGAAGGATCA GATCACGCAT CTTCCCGACA ACGCAGACCG
3851 TTCCGTGGCA AAGCAAAAGT TCAAAATCAC CAACTGGTCC CACCTGACCC
3901 CATGCCGAAC TCAGAAGTGA AACGCCGTAG CGCCGATGGT AGTGTGGGGT
3951 CTCCCCATGC GAGAGTAGGG AACTGCCAGG CATCAAATAA AACGAAAGGC
4001 TCAGTCGAAA GACTGGGCCT TTCGTTTTAT CTGTTGTTTG TCGGTGAACG
4051 CTCTCCTGAG TAGGACAAAT CCGCCGGGAG CGGATTTGAA CGTTGCGAAG
4101 CAACGGCCCG GAGGGTGGCG GGCAGGACGC CCGCCATAAA CTGCCAGGCA
4151 TCAAATTAAG CAGAAGGCCA TCCTGACGGA TGGCCTTTTT GCGTTTCTAC
4201 AAACTCTTTT GTTTATTTTT CTAAATACAT TCAAATATGT ATCCGCTCAT
4251 GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGTCTTCAAg
4301 aattcGAAAG TTAATGAATA GCACCGCCGg agctcggtac cCGGggatcc
4351 GCTTTACACT TTATGCTTCC GGCTCGTATA ATGTgtcgac CGAAAGTTAA
4401 TGAATAGCAC CCCgcgggccg cACACAGGAA ACAGCTATGA CCATGATTAC
4451 GAATTTCGAC ctgcagCCaa gcttGGCACT GGCCGTCGTT TTACAACGTC
4501 GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT
4551 CCCCCTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC
4601 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGCGCTTT GCCTGGTTTC
4651 CGGCACCAGA AGCGGTGCCG GAAGCTGGCT GGAGTGCGAT CTTCCTGAGG
4701 CCGATACTGT CGTCGTCCCC TCAAACTGGC AGATGCACGG TTACGATGCG

```
4751 CCCATCTACA CCAACGTAAC CTATCCCATT ACGGTCAATC CGCCGTTTGT
4801 TCCCAcggag aatccgacgg ggtgctattc attaactttc aatgttgatg
4851 AAAGCTGGCT ACAGGAAGGC CAGACGCGAA TTATTTTTGA TGGCGTTAAC
4901 TCGGCGTTTC ATCTGTGGTG CAACGGGCGC TGGGTCGGTT ACGGCCAGGA
4951 CAGTCGTTTG CCGTCTGAAT TTGACCTGAG CGCATTTTTA CGCGCCGGAG
5001 AAAACCGCCT CGCGGTGATG GTGCTGCGTT GGAGTGACGG CAGTTATCTG
5051 GAAGATCAGG ATATGTGGCG GATGAGCGGC ATTTTCCGTG ACGAATTTCT
5101 GCCATTCATC CGCTTATTAT CACTTATTCA GGCGTAGCAC CAGGCGTTTA
5151 AGGGCACCAA TAACTGCCTT AAAAAAATTA CGCCCCGCCC TGCCACTCAT
5201 CGCAGTACTG TTGTAATTCA TTAAGCATTC TGCCGACATG GAAGCCATCA
5251 CAGACGGCAT GATGAACCTG AATCGCCAGC GGCATCAGCA CCTTGTCGCC
5301 TTGCGTATAA TATTTGCCCA TGGTGAAAAC GGGGGCGAAG AAGTTGTCCA
5351 TATTGGCCAC GTTTAAATCA AAACTGGTGA AACTCACCCA GGGATTGGCT
5401 GAGACGAAAA ACATATTCTC AATAAACCCT TTAGGGAAAT AGGCCAGGTT
5451 TTCACCGTAA CACGCCACAT CTTGCGAATA TATGTGTAGA AACTGCCGGA
5501 AATCGTCGTG GTATTCACTC CAGAGCGATG AAAACGTTTC AGTTTGCTCA
5551 TGGAAAACGG TGTAACAAGG GTGAACACTA TCCCATATCA CCAGCTCACC
5601 GTCTTTCATT GCCATACG
```

- [proxWT-1] *Sal*I – *Not*I sequence insert in pJ1653
  ```
  4385 gtcgacCAAa tgactcttTC CCACAACCgc ggccgc
  ```

- [proxCREB-1] *Sal*I – *Not*I sequence insert in pJ1653
  ```
  4385 gtcgacCAat gacgtcatTC CCACAACCgc ggccgc
  ```

- [distWT] *Eco*RI – *Sac*I gene insert in pJ1653
  ```
  4300 gaattcAAat gactcttGTA CAAATTGCCG gagctc
  ```

- [distCREB] *Eco*RI – *Sac*I gene insert in pJ1653
  ```
  4300 gaattcAAat gacgtcatGT ACAATTGCCG gagctc
  ```

- [CC-85] *Eco*RI – *Not*I gene insert in pJ1653
  ```
  4300 gaattcAAat gacgtcatGT ACAATTGCCG gagctcggta ccCGGgatc
  4350 cGCTTTACAC TTTATGCTTC CGGCTCGTAT AATGTgtcga cCAatgacgt
  4400 catTCCCACA ACCgcggccg c
  ```

- [CW-85.5] *Eco*RI – *Not*I gene insert in pJ1653
  ```
  4300 gaattcAAat gacgtcatGT ACAATTGCCG gagctcggta ccCGGgatc
  4350 cGCTTTACAC TTTATGCTTC CGGCTCGTAT AATGTgtcga cCAAatgact
  4400 cttTCCCACA ACCgcggccg c
  ```

- [CW-138.5] *Eco*RI – *Not*I gene insert in pJ1653
  ```
  4300 gaattcAAat gacgtcatGT ACAATTGCCG gagctcGCTG CTAACAAAGC
  4350 CCGAAAGGAA GCTGAGTTGG CTGCTGCCAC CGCTGAGCAA TAACTAGCgg
  4400 atccGCTTTA CACTTTATGC TTCCGGCTCG TATAATGTgt cgacCAAatg
  4450 actcttTCCC ACAACCgcgg ccgc
  ```

- [CW-187.5] *Eco*RI – *Not*I gene insert in pJ1653

```
4300 gaattcAAat gacgtcatGT ACAATTGCCG gagctcGCTG CTAACAAAGC
4350 CCGAAAGGAA GCTGAGTTGG CTGCTGCCAC CGCTGAGCAA TAACTAGCAT
4400 AACCCCTTGG GGCCTCTAAA CGGGTCTTGA GGGGTTTTTT GCTGAAAgga
4450 tccGCTTTAC ACTTTATGCT TCCGGCTCGT ATAATGTgtc gacCAAatga
4500 ctcttTCCCA CAACCgcggc cgc
```

- [CW-237.5] *Eco*RI – *Not*I gene insert in pJ1653

```
4300 gaattcAAat gacgtcatGT ACAATTGCCG gagctcGCTG CTAACAAAGC
4350 CCGAAAGGAA GCTGAGTTGG CTGCTGCCAC CGCTGAGCAA TAACTAGCAT
4400 AACCCCTTGG GGCCTCTAAA CGGGTCTTGA GGGGTTTTTT GCTGAAAGGA
4450 GGAACTATAT CCGGATCTGG CGTAATAGCG AAGAGGCCCG CACCGATgga
4500 tccGCTTTAC ACTTTATGCT TCCGGCTCGT ATAATGTgtc gacCAAatga
4550 ctcttTCCCA CAACCgcggc cgc
```

- [proxCREB-2 and proxWT-2] *Eco*RI – *Sac*I linker sequence in proxCREB-1 and proxWT-1

```
4300 gaattcCAGA CGTAACAGCA CCACGGTGGT GGTGAACACG GTGGGCTCAG
4350 AGAATCCGGA TGAAGCCGGG CGTTACAGCA TGGATGTGGA GTACGGTCAG
4400 TACAGTGTCA TCCTGCAGGT TGACGGTTTT CCACCATCGC ACGCCGGGAC
4450 CATCACCGTG TATGAAGATT CACAACCGGG GACGCTGAAT GATTTTCTCT
4500 GTGCCATGAC GGAGGATGAT GCCCGGCCGG AGGTGCTGCG TCGTCTTGAA
4550 CTGATGGTGG AAGAGGTGGC GCGTAACGCG TCCGTGGTGG CACAGAGTAC
4600 GGCAGACGCG AAGAAATCAG CCGGCGATGC CAGTGCATCA GCTGCTCAGG
4650 TCGCGGCCCT TGTGACTGAT GCAACTGACT CAGCACGCGC CGCCAGCACG
4700 TCCGCCGGAC AGGCTGCATC GTCAGCTCAG GAAGCGTCCT CCGGCGCAGA
4750 AGCGGCATCA GCAAAGGCCA CTGAAGCGGA AAAAAGTGCC GCAGCCGCAG
4800 AGTCCTCAAA AAACGCGGCG GCCACCAGTG CCGGTGCGGC GAAAACGTCA
4850 GAAACGAATG CTGCAGCGTC ACAACAATCA GCCGCCACGT CTGCCTCCAC
4900 CGCGGCCACG AAAGCGTCAG AGGCCGCCAC TTCAGCACGA GATGCGGTGG
4950 CCTCAAAAGA GGCAGCAAAA TCATCAGAAA CGAACGCATC ATCAAGTGCC
5000 GGTCGTGCAG CTTCCTCGGC AACGGCGGCA GAAAATTCTG CCAGGGCGGC
5050 AAAAACGTCC GAGACGAATG CCAGGTCATC TGAAACAGCA GCGGAACGGA
5100 GCGCCTCTGC CGCGGCAGAC GCAAAAACAG CGGCGGCGGG GAGTGCGTCA
5150 ACGGCATCCA CGAAGGCGAC AGAGGCTGCG GGAAGTGCGG TATCAGCATC
5200 GCAGAGCAAA AGTGCGGCAG AAGCGGCGGC AATACGTGCA AAAAATTCGG
5250 CAAAACGTGC AGAAGATATA GCTTCAGCTG TCGCGCTTGA GGATGCGGAC
5300 ACAACGAGAA AGGGGATAGT GCAGCTCAGC AGTGCAACCA ACAGCACGTC
5350 TGAAACGCTT GCTGCAACGC CAAAGGCGGT TAAGGTGGTA ATGGgagctc
```

- [IC-300 and IW-300.5] *Eco*RI – *Sac*I gene insert in proxCREB-2 and proxWT-2

```
4300 gaattcCAGA CGTAACAGCA CCACGGTGGT GGTGAACACG GTGGGCTCAG
4350 AGAATCCGGA TGAAGCCGGG CGTTACAGCA TGGATGTGGA GTACGGTCAG
4400 TACAGTGTCA TCCTGCAGGT TGACGGTTTT CCACCATCGC ACGCCGGGAC
4450 CATCACCGTG TATGAAGATT CACAACCGGG GACGCTGAAT GATTTTCTCT
```

4500 GTGCCATGAC GGAGGATGAT GCCCGGCCGG AGGTGCTGCG TCGTCTTGAA
4550 CTGATGGTGG AAGAGGTGGC GCGTAACGCG TCCGTGGTGG CACAGAGTAC
4600 GGCAGACGCG AAGAAATCAG CCGGCGATGC CAGTGCATCA GCTGCTCAGG
4650 TCGCGGCCCT TGTGACTGAT GCAACTGACT CAGCACGCGC CGCCAGCACG
4700 TCCGCCGGAC AGGCTGCATC GTCAGCTCAG GAAGCGTCCT CCGGCGCAGA
4750 AGCGGCATCA GCAAAGGCCA CTGAAGCGGA AAAAAGTGCC GCAGCCGCAG
4800 AGTCCTCAAA AAACGCGGCG GCCACCAGTG CCGGTGCGGC GAAAACGTCA
4850 GAAACGAATG CTGCAGCGTC ACAACAATCA GCCGCCACGT CTGCCTCCAC
4900 CGCGGCCACG AAAGCGTCAG AGGCCGCCAC TTCAGCACGA GATGCGGTGG
4950 CCTCAAAAGA GGCAGCAAAA TCATCAGAAA CGAACGCATC ATCAAGTGCC
5000 GGTCGTGCAG CTTCCTCGGC AACGGCGGCA GAAAATTCTG CCAGGGCGGC
5050 AAAAACGTCC GAGACGAATG CCAGGTCATC TGAAACAGCA GCGGAACGGA
5100 GCGCCTCTGC CGCGGCAGAC GCAAAAACAG CGGCGGCGGG GAGTGCGTCa
5150 ctagtATgtc atatgacGAt gtacaCTGCG GGAAGTGCGG TATCAGCATC
5200 GCAGAGCAAA AGTGCGGCAG AAGCGGCGGC AATACGTGCA AAAAATTCGG
5250 CAAAACGTGC AGAAGATATA GCTTCAGCTG TCGCGCTTGA GGATGCGGAC
5300 ACAACGAGAA AGGGGATAGT GCAGCTCAGC AGTGCAACCA ACAGCACGTC
5350 TGAAACGCTT GCTGCAACGC CAAAGGCGGT TAAGGTGGTA ATGGgagctc

- [IC-450 and IW-450.5] *Eco*RI – *Sac*I gene insert in proxCREB-2 and proxWT-2

4300 gaattcCAGA CGTAACAGCA CCACGGTGGT GGTGAACACG GTGGGCTCAG
4350 AGAATCCGGA TGAAGCCGGG CGTTACAGCA TGGATGTGGA GTACGGTCAG
4400 TACAGTGTCA TCCTGCAGGT TGACGGTTTT CCACCATCGC ACGCCGGGAC
4450 CATCACCGTG TATGAAGATT CACAACCGGG GACGCTGAAT GATTTTCTCT
4500 GTGCCATGAC GGAGGATGAT GCCCGGCCGG AGGTGCTGCG TCGTCTTGAA
4550 CTGATGGTGG AAGAGGTGGC GCGTAACGCG TCCGTGGTGG CACAGAGTAC
4600 GGCAGACGCG AAGAAATCAG CCGGCGATGC CAGTGCATCA GCTGCTCAGG
4650 TCGCGGCCCT TGTGACTGAT GCAACTGACT CAGCACGCGC CGCCAGCACG
4700 TCCGCCGGAC AGGCTGCATC GTCAGCTCAG GAAGCGTCCT CCGGCGCAGA
4750 AGCGGCATCA GCAAAGGCCA CTGAAGCGGA AAAAAGTGCC GCAGCCGCAG
4800 AGTCCTCAAA AAACGCGGCG GCCACCAGTG CCGGTGCGGC GAAAACGTCA
4850 GAAACGAATG CTGCAGCGTC ACAACAATCA GCCGCCACGT CTGCCTCCAC
4900 CGCGGCCACG AAAGCGTCAG AGGCCGCCAC TTCAGCACGA GATGCGGTGG
4950 CCTCAAAAGA GGCAGCAAAA TCATCAGAAA CGAACGCATC ATCAAGTGCa
5000 ctagtTGgtc atatgacGGt gtacaCGGCA GAAAATTCTG CCAGGGCGGC
5050 AAAAACGTCC GAGACGAATG CCAGGTCATC TGAAACAGCA GCGGAACGGA
5100 GCGCCTCTGC CGCGGCAGAC GCAAAAACAG CGGCGGCGGG GAGTGCGTCA
5150 ACGGCATCCA CGAAGGCGAC AGAGGCTGCG GGAAGTGCGG TATCAGCATC
5200 GCAGAGCAAA AGTGCGGCAG AAGCGGCGGC AATACGTGCA AAAAATTCGG
5250 CAAAACGTGC AGAAGATATA GCTTCAGCTG TCGCGCTTGA GGATGCGGAC
5300 ACAACGAGAA AGGGGATAGT GCAGCTCAGC AGTGCAACCA ACAGCACGTC
5350 TGAAACGCTT GCTGCAACGC CAAAGGCGGT TAAGGTGGTA ATGGgagctc

- [IC-650 and IW-650.5] *Eco*RI – *Sac*I gene insert in proxCREB-2 and proxWT-

2

```
4300 gaattcCAGA CGTAACAGCA CCACGGTGGT GGTGAACACG GTGGGCTCAG
4350 AGAATCCGGA TGAAGCCGGG CGTTACAGCA TGGATGTGGA GTACGGTCAG
4400 TACAGTGTCA TCCTGCAGGT TGACGGTTTT CCACCATCGC ACGCCGGGAC
4450 CATCACCGTG TATGAAGATT CACAACCGGG GACGCTGAAT GATTTTCTCT
4500 GTGCCATGAC GGAGGATGAT GCCCGGCCGG AGGTGCTGCG TCGTCTTGAA
4550 CTGATGGTGG AAGAGGTGGC GCGTAACGCG TCCGTGGTGG CACAGAGTAC
4600 GGCAGACGCG AAGAAATCAG CCGGCGATGC CAGTGCATCA GCTGCTCAGG
4650 TCGCGGCCCT TGTGACTGAT GCAACTGACT CAGCACGCGC CGCCAGCACG
4700 TCCGCCGGAC AGGCTGCATC GTCAGCTCAG GAAGCGTCCT CCGGCGCAGA
4750 AGCGGCATCA GCAAAGGCCA CTGAAGCGGA AAAAAGTGCC GCAGCCGCAa
4800 ctagtTCgtc atatgacGCt gtacaCAGTG CCGGTGCGGC GAAAACGTCA
4850 GAAACGAATG CTGCAGCGTC ACAACAATCA GCCGCCACGT CTGCCTCCAC
4900 CGCGGCCACG AAAGCGTCAG AGGCCGCCAC TTCAGCACGA GATGCGGTGG
4950 CCTCAAAAGA GGCAGCAAAA TCATCAGAAA CGAACGCATC ATCAAGTGCC
5000 GGTCGTGCAG CTTCCTCGGC AACGGCGGCA GAAAATTCTG CCAGGGCGGC
5050 AAAAACGTCC GAGACGAATG CCAGGTCATC TGAAACAGCA GCGGAACGGA
5100 GCGCCTCTGC CGCGGCAGAC GCAAAAACAG CGGCGGCGGG GAGTGCGTCA
5150 ACGGCATCCA CGAAGGCGAC AGAGGCTGCG GGAAGTGCGG TATCAGCATC
5200 GCAGAGCAAA AGTGCGGCAG AAGCGGCGGC AATACGTGCA AAAAATTCGG
5250 CAAAACGTGC AGAAGATATA GCTTCAGCTG TCGCGCTTGA GGATGCGGAC
5300 ACAACGAGAA AGGGGATAGT GCAGCTCAGC AGTGCAACCA ACAGCACGTC
5350 TGAAACGCTT GCTGCAACGC CAAAGGCGGT TAAGGTGGTA ATGGgagctc
```

- [IC-900 and IW-900.5] *Eco*RI – *Sac*I gene insert in proxCREB-2 and proxWT-2

```
4300 gaattcCAGA CGTAACAGCA CCACGGTGGT GGTGAACACG GTGGGCTCAG
4350 AGAATCCGGA TGAAGCCGGG CGTTACAGCA TGGATGTGGA GTACGGTCAG
4400 TACAGTGTCA TCCTGCAGGT TGACGGTTTT CCACCATCGC ACGCCGGGAC
4450 CATCACCGTG TATGAAGATT CACAACCGGG GACGCTGAAT GATTTTCTCT
4500 GTGCCATGAC GGAGGATGAT GCCCGGCCGG AGGTGCTGCG TCGTCTTGAa
4550 ctagtGGgtc atatgacGGt gtacaACGCG TCCGTGGTGG CACAGAGTAC
4600 GGCAGACGCG AAGAAATCAG CCGGCGATGC CAGTGCATCA GCTGCTCAGG
4650 TCGCGGCCCT TGTGACTGAT GCAACTGACT CAGCACGCGC CGCCAGCACG
4700 TCCGCCGGAC AGGCTGCATC GTCAGCTCAG GAAGCGTCCT CCGGCGCAGA
4750 AGCGGCATCA GCAAAGGCCA CTGAAGCGGA AAAAAGTGCC GCAGCCGCAG
4800 AGTCCTCAAA AAACGCGGCG GCCACCAGTG CCGGTGCGGC GAAAACGTCA
4850 GAAACGAATG CTGCAGCGTC ACAACAATCA GCCGCCACGT CTGCCTCCAC
4900 CGCGGCCACG AAAGCGTCAG AGGCCGCCAC TTCAGCACGA GATGCGGTGG
4950 CCTCAAAAGA GGCAGCAAAA TCATCAGAAA CGAACGCATC ATCAAGTGCC
5000 GGTCGTGCAG CTTCCTCGGC AACGGCGGCA GAAAATTCTG CCAGGGCGGC
5050 AAAAACGTCC GAGACGAATG CCAGGTCATC TGAAACAGCA GCGGAACGGA
5100 GCGCCTCTGC CGCGGCAGAC GCAAAAACAG CGGCGGCGGG GAGTGCGTCA
5150 ACGGCATCCA CGAAGGCGAC AGAGGCTGCG GGAAGTGCGG TATCAGCATC
5200 GCAGAGCAAA AGTGCGGCAG AAGCGGCGGC AATACGTGCA AAAAATTCGG
```

```
5250 CAAAACGTGC AGAAGATATA GCTTCAGCTG TCGCGCTTGA GGATGCGGAC
5300 ACAACGAGAA AGGGGATAGT GCAGCTCAGC AGTGCAACCA ACAGCACGTC
5350 TGAAACGCTT GCTGCAACGC CAAAGGCGGT TAAGGTGGTA ATGGgagctc
```

- [IC-1150 and IW-1150.5] *Eco*RI – *Sac*I gene insert in proxCREB-2 and proxWT-2

```
4300 gaattcCgtc atatgacGCt gtacaGTGGT GGTGAACACG GTGGGCTCAG
4350 AGAATCCGGA TGAAGCCGGG CGTTACAGCA TGGATGTGGA GTACGGTCAG
4400 TACAGTGTCA TCCTGCAGGT TGACGGTTTT CCACCATCGC ACGCCGGGAC
4450 CATCACCGTG TATGAAGATT CACAACCGGG GACGCTGAAT GATTTTCTCT
4500 GTGCCATGAC GGAGGATGAT GCCCGGCCGG AGGTGCTGCG TCGTCTTGAA
4550 CTGATGGTGG AAGAGGTGGC GCGTAACGCG TCCGTGGTGG CACAGAGTAC
4600 GGCAGACGCG AAGAAATCAG CCGGCGATGC CAGTGCATCA GCTGCTCAGG
4650 TCGCGGCCCT TGTGACTGAT GCAACTGACT CAGCACGCGC CGCCAGCACG
4700 TCCGCCGGAC AGGCTGCATC GTCAGCTCAG GAAGCGTCCT CCGGCGCAGA
4750 AGCGGCATCA GCAAAGGCCA CTGAAGCGGA AAAAAGTGCC GCAGCCGCAG
4800 AGTCCTCAAA AAACGCGGCG GCCACCAGTG CCGGTGCGGC GAAAACGTCA
4850 GAAACGAATG CTGCAGCGTC ACAACAATCA GCCGCCACGT CTGCCTCCAC
4900 CGCGGCCACG AAAGCGTCAG AGGCCGCCAC TTCAGCACGA GATGCGGTGG
4950 CCTCAAAAGA GGCAGCAAAA TCATCAGAAA CGAACGCATC ATCAAGTGCC
5000 GGTCGTGCAG CTTCCTCGGC AACGGCGGCA GAAAATTCTG CCAGGGCGGC
5050 AAAAACGTCC GAGACGAATG CCAGGTCATC TGAAACAGCA GCGGAACGGA
5100 GCGCCTCTGC CGCGGCAGAC GCAAAAACAG CGGCGGCGGG GAGTGCGTCA
5150 ACGGCATCCA CGAAGGCGAC AGAGGCTGCG GGAAGTGCGG TATCAGCATC
5200 GCAGAGCAAA AGTGCGGCAG AAGCGGCGGC AATACGTGCA AAAAATTCGG
5250 CAAAACGTGC AGAAGATATA GCTTCAGCTG TCGCGCTTGA GGATGCGGAC
5300 ACAACGAGAA AGGGGATAGT GCAGCTCAGC AGTGCAACCA ACAGCACGTC
5350 TGAAACGCTT GCTGCAACGC CAAAGGCGGT TAAGGTGGTA ATGGgagctc
```

2.4 Translated protein sequences

His-tagged recombinant proteins in pRSETA and pFastBac1

- [empty pRSETA - 71 aa]
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSELEICSWYH
GIRSLIRLLTKPERKLSWLLPPLSNN

- [4har - 117 aa]
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSDPAALKRA
RNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVA
RLARQVRALADSLMQLARQVSRLADSL

- [4hee - 117 aa]
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSDPAALKRA
RNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVEE
LARQVRALADSLMQLARQVSRLADSL

- [lzar - 117 aa]
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSDPAALKRA
RNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVA
RLKKLVRALADSLMQLARQVSRLESGQ

- [lzee - 117 aa]
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSDPAALKRA
RNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVEE
LLSKVRALADSLMQLARQVSRLESGQ

- [LZD73 in pRSETA - 119 aa]
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSDPAALKRA
RNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVG
ELQKLQRVKRARNTEAARRSRARKAALKG

- [LZD80 in pRSETA - 126 aa]
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSDPAALKRA
RNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVEE
LLSKVGELQKLQRVKRARNTEAARRSRARKAALKG

- [LZD87 in pRSETA - 133 aa]
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSDPAALKRA
RNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVEE
LLSKVRALADSLGELQKLQRVKRARNTEAARRSRARKAALKG

GST-tagged recombinant proteins in pGEX6P1 and pGEX6P1b

- [empty pGEX6P1 - 244 aa]
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLE
GAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLN
GDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK
YLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSPEFPGRL
ERPHRD

- [4har in pGEX6P1 - 312 aa]
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLE
GAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLN
GDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK
YLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSDPAALKR
ARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLV
ARLARQVRALADSLMQLARQVSRLADSL

- [lzee in pGEX6P1 - 312 aa]
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLE

GAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLN
GDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK
YLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSDPAALKR
ARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVE
ELLSKVSALADSLMQLARQVSRLESGQ

- [lzee.v2 in pGEX6P1 - 312 aa]
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLE
GAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLN
GDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK
YLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSDPAALKR
ARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVE
ELLSKVRALADSLMQLARQVSRLESGQ

- [lzee.v2 libctrl in pGEX6P1 - 312 aa]
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLE
GAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLN
GDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK
YLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSDPAALKR
ARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVE
ELLSKLRALADSLMQLARQVSRLESGQ

- [LZD73 in pGEX6P1 - 314 aa]
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLE
GAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLN
GDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK
YLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSDPAALKR
ARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLV
GELQKLQRVKRARNTEAARRSRARKAALKG

- [LZD80 in pGEX6P1 - 321 aa]
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLE
GAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLN
GDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK
YLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSDPAALKR
ARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVE
ELLSKVGELQKLQRVKRARNTEAARRSRARKAALKG

- [LZD87 in pGEX6P1 - 328 aa]
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLE
GAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLN

GDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK
YLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSDPAALKR
ARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVE
ELLSKVRALADSLGELQKLQRVKRARNTEAARRSRARKAALKG

Untagged recombinant proteins in pBAD/Myc-HisA

- [empty pBAD-Myc-HisA - 30 aa]
MDPSSRSAAGTIWEFEAWARTKTHLRRGSE

- [lzee in pBAD-Myc-HisA - 102 aa]
MDPSSRSAAGTEVLFQGPLGSDPAALKRARNTEAARRSRARKLQRMK
QLEDKVEELLSKNYHLENEVARLKKLVEELLSKVRALADSLMQLARQ
VSRLESGQ

- [LZD73 in pBAD-Myc-HisA - 104 aa]
MDPSSRSAAGTEVLFQGPLGSDPAALKRARNTEAARRSRARKLQRMK
QLEDKVEELLSKNYHLENEVARLKKLVGELQKLQRVKRARNTEAARR
SRARKAALKG

- [LZD80 in pBAD-Myc-HisA - 111 aa]
MDPSSRSAAGTEVLFQGPLGSDPAALKRARNTEAARRSRARKLQRMK
QLEDKVEELLSKNYHLENEVARLKKLVEELLSKVGELQKLQRVKRAR
NTEAARRSRARKAALKG

- [LZD87 in pBAD-Myc-HisA - 118 aa]
MDPSSRSAAGTEVLFQGPLGSDPAALKRARNTEAARRSRARKLQRMK
QLEDKVEELLSKNYHLENEVARLKKLVEELLSKVRALADSLGELQKLQ
RVKRARNTEAARRSRARKAALKG

2.5 PCC qPCR primers

- FPnorm2-PCC (FPnorm)
  `1 CAGGCGTTTA AGGGCACCAA TA`

- RPnorm2-PCC (RPnorm)
  `1 GCTGGCGATT CAGGTTCATC AT`

- qFPctrl1-HinP1I-428 (FPctrl1)
  `1 CATTTTCGCC AAAAGTTGGC CCAGGG`

- qFPctrl2-HinP1I-581 (FPctrl2)
  `1 CTATCAGCTG TCCCTCCTGT TCAGC`

- qRPctrl1-HinP1I-428 (RPctrl1)
  `1 GAAACAGAAG CCACTGGAGC ACCTC`

- qRPctrl2-HinP1I-581 (RPctrl2)
  `1 CTTCACTGAC ACCCTCATCA GTGCC`
- qFP-HinP1I-5676 (A1)
  `1 AAAACCCTGG CGTTACCCAA CTTAATCG`
- qRP-HinP1I-5282 (A2)
  `1 CCATTACCAC CTTAACCGCC TTTGG`
- qFP-HinP1I-5282 (B1)
  `1 GGTATCAGCA TCGCAGAGCA AAAGTG`
- qRP-HinP1I-5100 (B2)
  `1 CACTTTTGCT CTGCGATGCT GATACC`
- qFP-HinP1I-5100 (C1)
  `1 GCAGCAAAAT CATCAGAAAC GAACGCATC`
- qRP-HinP1I-4743 (C2)
  `1 AGCATTCGTT TCTGACGTTT TCGCC`
- qFP-HinP1I-4686 (D1)
  `1 AGTACGGCAG ACGCGAAGAA ATCAG`
- qRP-HinP1I-3930 (D2, old)
  `1 CCGACAAACA ACAGATAAAA CGAAAGGCC`
- qRP-HinP1I-3930 (D2, new)
  `1 CCTACTCAGG AGAGCGTTCA CCG`

Bibliography

Becker, N. A., Kahn, J. D., & Maher, L. J. (2005). Bacterial repression loops require enhanced DNA flexibility. *Journal of molecular biology*, *349*(4), 716–730.

Becker, N. A., Kahn, J. D., & Maher, L. J. (2007). Effects of nucleoid proteins on DNA repression loop formation in *Escherichia coli*. *Nucleic acids research*, *35*(12), 3988–4000.

Bellomy, G. R., Mossing, M. C., & Record, M. T. (1988). Physical properties of DNA *in vivo* as probed by the length dependence of the *lac* operator looping process. *Biochemistry*, *27*(11), 3900–3906.

Cloutier, T. E., & Widom, J. (2004). Spontaneous sharp bending of double-stranded dna. *Molecular cell*, *14*, 355–62.

Cloutier, T. E., & Widom, J. (2005). Dna twisting flexibility and the formation of sharply looped protein-dna complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(10), 3645–3650.

Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nature methods*, *3*(1), 17–21.

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science (New York, N.Y.)*, *295*(5558), 1306–1311.

Du, Q., Smith, C., Shiffeldrim, N., Vologodskaia, M., & Vologodskii, A. (2005). Cyclization of short dna fragments and bending fluctuations of the double helix. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 5397–402.

Echols, H. (1990). Nucleoprotein structures initiating DNA replication, transcription, and site-specific recombination. *Journal of Biological Chemistry*, *265*, 14697–14700.

Friedman, A. M., Fischmann, T. O., & Steitz, T. A. (1995). Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science (New York, N.Y.)*, *268*(5218), 1721–1727.

Glick, B. R. (1995). Metabolic load and heterologous gene expression. *Biotechnology advances*, *13*(2), 247–261.

Goodson, K. A., Wang, Z., Haeusler, A. R., Kahn, J. D., & English, D. S. (2013). LacI-DNA-IPTG loops: equilibria among conformations by single-molecule FRET. *Journal of Physical Chemistry B*, *117*, 4713–4722.

Gowetski, D. B., Kodis, E. J., & Kahn, J. D. (2013). Rationally designed coiled-coil DNA looping peptides control DNA topology. *Nucleic acids research*, *41*(17), 8253–8265.

Haeusler, A. R., Goodson, K. A., Lillian, T. D., Wang, X., Goyal, S., Perkins, N. C., & Kahn, J. D. (2012). FRET studies of a landscape of lac repressor-mediated DNA loops. *Nucleic acids research*, *40*, 4432–4445.

Hayes, J. D., Flanagan, J. U., & Jowsey, I. R. (2005). Glutathione transferases. *Annual Review of Pharmacology and Toxicology*, *45*, 51–88.

Hill, D., Hope, I., Macke, J., & Struhl, K. (1986). Saturation mutagenesis of the yeast *his3* regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. *Science*, *234*(4775), 451–457.

Hollenbeck, J. J., Gurnon, D. G., Fazio, G. C., Carlson, J. J., & Oakley, M. G. (2001). A GCN4 variant with a C-terminal basic region binds to DNA with wild-type affinity. *Biochemistry*, *40*, 13833–13839.

Ji, X., Zhang, P., Armstrong, R. N., & Gilliland, G. L. (1992). The three-dimensional structure of a glutathione S-transferase from the mu gene class. structural

analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-Å resolution. *Biochemistry*, *31*(42), 10169–10184.

Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., & Hecht, M. H. (1993). Protein design by binary patterning of polar and nonpolar amino acids. *Science (New York, N.Y.)*, *262*(5140), 1680–1685.

Keller, W., König, P., & Richmond, T. J. (1995). Crystal structure of a bZIP/DNA complex at 2.2 Å: determinants of DNA specific recognition. *Journal of molecular biology*, *254*(4), 657–667.

König, P., & Richmond, T. J. (1993). The x-ray structure of the gcn4-bzip bound to atf/creb site dna shows the complex depends on dna flexibility. *Journal of molecular biology*, *233*, 139–54.

Krämer, H., Niemöller, M., Amouyal, M., Revet, B., von Wilcken-Bergmann, B., & Müller-Hill, B. (1987). lac repressor forms loops with linear dna carrying two suitably spaced lac operators. *The EMBO journal*, *6*(5), 1481–1491.

Kratky, O., & Porod, G. (1949). Röntgenuntersuchung gelöster fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas*, *68*(12), 1106–1122.

Lilja, A. E., Jenssen, J. R., & Kahn, J. D. (2004). Geometric and dynamic requirements for DNA looping, wrapping and unwrapping in the activation of *E. coli glnAp2* transcription by NtrC. *Journal of molecular biology*, *342*(2), 467–478.

Lu, K., Ye, W., Zhou, L., Collins, L. B., Chen, X., Gold, A., ... Swenberg, J. A. (2010). Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers. *Journal of the American Chemical Society*, *132*(10), 3388–3399.

MacConkey, A. T. (1908). Bile salt media and their advantages in some bacteriological examinations. *The journal of hygiene*, *8*(3), 322–334.

Maru, Y., Afar, D. E., Witte, O. N., & Shibuya, M. (1996). The dimerization

property of glutathione S-transferase partially reactivates bcr-abl lacking the oligomerization domain. *Journal of Biological Chemistry*, *271*(26), 15353–15357.

Matthews, K. S. (1992). DNA looping. *Microbiological reviews*, *56*(1), 123–136.

Miele, A., Gheldof, N., Tabuchi, T. M., Dostie, J., & Dekker, J. (2006). Mapping chromatin interactions by chromosome conformation capture. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, Unit 21.11.

Miller, J. H. (1972). *Assay of beta-galactosidase* (J. H. Miller, Ed.). Cold Spring Harbor, NY: CSH Laboratory Press.

Mossing, M. C., & Record, M. T. (1986). Upstream operators enhance repression of the *lac* promoter. *Science (New York, N.Y.)*, *233*(4766), 889–892.

Murphy, C. I., Piwnica-Worms, H., Grünwald, S., Romanow, W. G., Francis, N., & Fan, H.-Y. (2004). Overview of the baculovirus expression system. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, Unit 16.9.

Oehler, S., Eismann, E. R., Krämer, H., & Müller-Hill, B. (1990). The three operators of the *lac* operon cooperate in repression. *The EMBO journal*, *9*(4), 973–979.

Parker, M. W., Lo Bello, M., & Federici, G. (1990). Crystallization of glutathione S-transferase from human placenta. *Journal of molecular biology*, *213*(2), 221–222.

Peters, J. P., Becker, N. A., Rueter, E. M., Bajzer, Z., Kahn, J. D., & Maher, L. J. (2011). Quantitative methods for measuring DNA flexibility *in vitro* and *in vivo*. *Methods in enzymology*, *488*, 287–335.

Rutkauskas, D., Zhan, H., Matthews, K. S., Pavone, F. S., & Vanzi, F. (2009). Tetramer opening in LacI-mediated DNA looping. *Proceedings of the National*

Academy of Sciences of the United States of America, *106*(39), 16627–16632.

Sasse-Dwight, S., & Gralla, J. D. (1988). Probing the *Escherichia coli glnALG* upstream activation mechanism *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America*, *85*, 8934–8938.

Schägger, H., & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical biochemistry*, *166*(2), 368–379.

Schleif, R. (2000). Regulation of the L-arabinose operon of *Escherichia coli*. *Trends in Genetics*, *16*, 559–565.

Shimada, J., & Yamakawa, H. (1984). Ring-closure probabilities for twisted wormlike chains. application to dna. *Macromolecules*, *17*, 689–698.

Shore, D., Langowski, J., & Baldwin, R. L. (1981). DNA flexibility studied by covalent closure of short fragments into circles. *Proceedings of the National Academy of Sciences of the United States of America*, *78*, 4833–4837.

Shuman, H. A., & Silhavy, T. J. (2003). The art and design of genetic screens: *Escherichia coli*. *Nature reviews. Genetics*, *4*(6), 419–431.

Smith, D. B., & Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*, *67*, 31–40.

Strauss-Soukup, J. K., & Maher, L. J. (1997). Dna bending by gcn4 mutants bearing cationic residues. *Biochemistry*, *36*, 10026–32.

Vanzi, F., Broggio, C., Sacconi, L., & Pavone, F. S. (2006). Lac repressor hinge flexibility and DNA looping: single molecule kinetics by tethered particle motion. *Nucleic acids research*, *34*(12), 3409–3420.

Weiss, D. S., Batut, J., Klose, K. E., Keener, J., & Kustu, S. (1991). The phospho-

rylated form of the enhancer-binding protein NTRC has an ATPase activity that is essential for activation of transcription. *Cell*, *67*, 155–167.

Whipple, F. W. (1998). Genetic analysis of prokaryotic and eukaryotic DNA-binding proteins in *Escherichia coli*. *Nucleic acids research*, *26*(16), 3700–3706.

Wiggins, P. A., Phillips, R., & Nelson, P. C. (2005). Exact theory of kinkable elastic polymers. *Physical review. E, Statistical, nonlinear, and soft matter physics*, *71*, 021909.

Zhang, Y., & Crothers, D. M. (2003). Statistical mechanics of sequence-dependent circular dna and its application for dna cyclization. *Biophysical journal*, *84*(1), 136–153.